



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS

Automatización del Sistema de Información aplicando Machine Learning en el
área logística en Villa Chicken S.A.C

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE: Ingeniero de sistemas

AUTORES:

Yoel Kereli Coronel Araujo (ORCID: 0000-0002-3968-4273)

Walther Feliciano Cáceres Guerra (ORCID: 0000-0002-9260-2273)

ASESOR:

Mg. Milner David Liendo Arévalo (0000-0002-7665-361X)

LÍNEA DE INVESTIGACIÓN:

Sistemas de información y comunicación

LIMA – PERÚ

2019

Dedicatoria

Agradecemos a Dios por la vida, nuestros padres que nos dieron las fuerzas para llevar adelante esta investigación, a nuestros hermanos que nos dieron las fortalezas necesarias. Asimismo, a nuestros amigos que confiaron en nosotros.

Caceres Walther y Coronel Yoel

Agradecimientos

Agradezco a dios por brindarnos el bienestar y salud, a nuestros padres, hermanos y amigos por apoyarnos en esta trayectoria. Asimismo, a nuestros asesores por brindarnos los conocimientos necesarios para llevar a cabo esta tesis.

Caceres Walther y Coronel Yoel

Página de Jurado

Declaratoria de autenticidad

Nosotros Walther Feliciano Cáceres Guerra con DNI N° 48757938, y Yoel Kereli Coronel Araujo con DNI N° 72024497 a efecto de cumplir con las disposiciones vigentes consideradas en el Reglamento de Grados y Títulos de la Universidad César Vallejo, Facultad de Ingeniería, Escuela Profesional de Ingeniería de Sistemas, declaro bajo juramento que toda la documentación que acompaño es veraz y auténtica. Asimismo, declaro también bajo juramento que todos los datos e información que se presenta en la presente tesis son auténticos y veraces. En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada por lo cual me someto a lo dispuesto en las normas académicas de la Universidad César Vallejo.

Lima, 6 de Julio de 2019



Walther Feliciano Cáceres Guerra
DNI: 48757938



Yoel Kereli Coronel Araujo
DNI: 72024497

Presentación

Señores miembros del jurado, en cumplimiento del Reglamento de Grados y Títulos de la Universidad César Vallejo presento ante ustedes la tesis titulada "Automatización del Sistema de Información aplicando Machine Learning en el área logística en Villa Chicken S.A.C", cuyo objetivo fue determinar la influencia de la aplicación de sistema de información utilizando Machine Learning en el área logística en Villa Chicken S.A.C. y que someto a vuestra consideración y espero que cumpla con los requisitos de aprobación para obtener el título profesional de ingeniero de sistemas. La investigación consta de seis capítulos. En el primer capítulo se explica la realidad problemática de varias empresas que en la actualidad no han podido prever el crecimiento de sus ventas o cual es el momento más adecuado para incrementarla, bajo esta problemática Villa Chicken S.A.C. ha presentado un crecimiento en su rubro, pero no ha logrado el incremento deseado. Por otro lado vemos investigaciones de diversos autores que analizan también esta problemática, y que han dado una solución a este problema de diversas maneras. Aplicando el método científico; definimos la formulación del problema como las hipótesis y objetivos; en el segundo capítulo se muestra definimos tanto el diseño de la investigación como las variables y las técnicas e instrumentos a emplear, en el tercer capítulo se detalla los datos evaluados estadísticamente, empleando para tal la estadística descriptiva como la inferencial, para la comprobación de las hipótesis planteadas en el segundo capítulo. En el cuarto capítulo se explica se realiza la discusión de los resultados obtenidos con otros autores que empleamos en el primer capítulo como base para esta investigación. En el quinto capítulo se presenta las conclusiones que obtuvimos y los resultados de la investigación. En el sexto capítulo se detalla finalmente describimos aquellas ideas que serían complementarias o mejoras en nuestra investigación.



Walther Feliciano Cáceres Guerra



Yoel Kereli Coronel Araujo

Índice

Dedicatoria.....	ii
Agradecimientos.....	iii
Página de Jurado.....	iv
Declaratoria de autenticidad	v
Presentación.....	vi
Resumen	xi
Abstract.....	xii
I. INTRODUCCIÓN.....	1
1.1 Realidad problemática.....	2
1.2 Trabajos previos.....	3
1.3 Teorías relacionadas al tema.....	11
1.4 Formulación del problema	30
1.5 Justificación del estudio	30
1.6 Hipótesis	32
1.7 Objetivos	32
II. MÉTODO.....	34
2.1 Diseño de la investigación.....	35
2.2 Variables, operacionalización	36
2.3 Población y muestra.....	37
2.4 Técnicas e instrumentos de recolección de datos, validez y confiabilidad	38
2.5 Métodos de análisis de datos	39
2.6 Aspectos éticos	39
III. RESULTADOS.....	40
3.1 Pruebas de normalidad.....	41
3.2 Dimensión 1: Inventario	41
3.3 Dimensión 2: Compras.....	46
3.4 Dimensión 3: Ventas.....	51
IV. DISCUSIÓN.....	56
V. CONCLUSIONES.....	59
VI. RECOMENDACIONES.....	61
VII. REFERENCIAS	63
ANEXOS	69

Índice de tablas

Tabla 1: Cálculo Estadístico descriptivos - Rotación de mercancías	42
Tabla 2: Tabla de Frecuencia Pre Test - Rotación de mercancías	42
Tabla 3 : Tabla de Frecuencias Post Test - Rotación de mercancías	43
Tabla 4: Prueba de Kolmogorov-Smirnow	44
Tabla 5: Prueba de hipótesis – rotación de mercancías.....	45
Tabla 6: Estadística de Prueba Wilcoxon.....	45
Tabla 7: Cálculos Estadísticos Descriptivos - Volumen de compras.....	47
Tabla 8: Tabla de Frecuencia Pre Test - Volumen de compras.....	47
Tabla 9 : Tabla de Frecuencia Post Test - Volumen de compras	48
Tabla 10 : Prueba de Kolmogorov- Volumen de compras	49
Tabla 11 Aplicación de la muestra relacionadas T student - Volumen de compras.....	50
Tabla 12: Cálculo de Datos descriptivos - Crecimiento de ventas.....	51
Tabla 13: Tabla de Frecuencia Pre Test - Crecimiento de ventas	52
Tabla 14: Tabla de Frecuencia Post Test - Crecimiento de ventas.....	52
Tabla 15: Prueba de Kolmogorov - Crecimiento de ventas	54
Tabla 16: Aplicación de la muestra relacionadas T student - Crecimiento de Ventas	55
Tabla 17 Ficha de registros- rotación de mercancías (Pre -Test)	71
Tabla 18 Ficha de registros- volumen de compras (Pre Test)	73
Tabla 19 Ficha de registros- Crecimiento de Ventas (Pre Test).....	75
Tabla 20 Ficha de registros- rotación de mercancías (Post -Test).....	77
Tabla 21 Ficha de registro- volumen de compras (Post Test)	79
Tabla 22 Ficha de registro- Crecimiento de Ventas (Post Test).....	81
Tabla 23 Serie de Tiempo Rotación de Mercancías(POLLO S/M GRANEL B4) Anuales por Fecha Festiva - Preparación de los Datos.....	87
Tabla 24 Serie de Tiempo Rotación de Mercancías(Papa Pre frita AVIKO Caja x 10 Kg) Anuales por Fecha Festiva - Preparación de los Datos	88
Tabla 25 Serie de Tiempo Rotación de Mercancías(Churrasco (200gr)) Anuales por Fecha Festiva - Preparación de los Datos	88
Tabla 26 Serie de Tiempo Rotación de Mercancías POLLO S/M GRANEL B4)- Mensual (Mayo) - Preparación de los Datos.....	88
Tabla 27 Serie de Tiempo Rotación de Mercancías POLLO S/M GRANEL B4)- Mensual (Junio) - Preparación de los Datos	89
Tabla 28 Evaluación de Resultados Objetivo 1 desde el Sistema de Predicción Logístico	91
Tabla 29 Evaluación de Resultados Objetivo 2 desde el Sistema de Predicción Logístico	91

Índice de figuras

Figura 1 : Esquema de los 4 niveles CRISP-DM	13
Figura 2: Fases de CRISP-DM.....	13
Figura 3 : Fases de la metodología SEMMA	20
Figura 4: Fases de la metodología KDD	21
Figura 5: Escenarios de la metodología KD	26
Figura 6 Correspondencia entre las fases de cada metodología	28
Figura 7. Cuadro de Operacionalización de las variables	36
Figura 8 Puntajes obtenidos Pre y Test, Post Test	41
Figura 9 : Gráfico de puntajes obtenidos en el Pre Test- Rotación de mercancías	43
Figura 10: Gráfico de puntajes obtenidos en el Post Test- Rotación de Mercancías	44
Figura 11 Puntajes Obtenidos Pre Test, Post Test y diferencia- Volumen de compras	46
Figura 12 Gráfico de puntajes obtenidos en el Pre Test - Volumen de compras	48
Figura 13: Gráfico de puntajes obtenidos en el Post Test - Volumen de compras.....	49
Figura 14: Puntajes obtenidos en el Pre Test, Post Test y la diferencia - Crecimiento de ventas	51
Figura 15: Gráfico de puntajes obtenidos en el Pre Test – Crecimiento de Ventas	53
Figura 16: Gráfico de puntajes obtenidos en el Post Test - Crecimiento de ventas	53
Figura 17: Matriz de Consistencia - Automatización de sistema de información aplicando machine learning en el área logística de Villa Chicken S.A.C.....	70
Figura 18: Recolección de Datos (Movimiento de inventario por Fecha Festiva periodo 2016 al 2018)- Fase Inicial.....	84
Figura 19 Recolección de Datos (Movimiento de inventario Mensual(Mayo-Junio) Festiva periodo 2016 al 2018)- Fase Inicial.....	85
Figura 20 Script Sql Rotación de mercancías - Preparación de los Datos	87
Figura 21: Modelamiento del clasificador desde el Sistema de Predicción Logístico- Valores de entrenamiento	90
Figura 22 : Sistema de predicción Logístico- Inicio Sesión.....	93
Figura 23: Sistema de predicción Logístico – Predicción cantidad de rotación POLLO S/M GGRANEL B4- Mensual	94
Figura 24: Sistema de predicción Logístico – Predicción cantidad rotación POLLO S/M GGRANEL B4- MENSUAL – Regresión Lineal	94
Figura 25: Sistema de predicción Logístico - Predicción cantidad rotación POLLO S/M GGRANEL B4 2019 – Día del Pollo-Regresión Lineal	95
Figura 26: Sistema de predicción Logístico – Predicción de cantidad de rotación POLLO S/M GGRANEL B4 2019 FECHA FESTIVO -DÍA DE LA MADRE-Regresión Lineal.....	95
Figura 27: Estructura del aplicativo web- Sistema predicción logística.	96
Figura 28: Sistema de predicción Logística – Capa Contoller	97
Figura 29: Sistema de Predicción Logística – Capa Model	97
Figura 30 Modelo Físico de la Base De datos.....	98

Índice de anexos

Anexo 1: Matriz de consistencia	70
Anexo 2: Instrumento de recolección de datos	71
Anexo 3: Desarrollo	83
Anexo 4: Permiso de uso de datos de institución.....	99
Anexo 5: Acta de aprobación de Originalidad de Tesis.....	100
Anexo 6: Pantallazo de Software Turnitin	101
Anexo 7: Autorización de publicación de tesis.....	102
Anexo 8: Autorización de la versión final del trabajo de investigación.....	103

Resumen

La presente tesis muestra la comprensión del negocio, preparación, modelamiento y evaluación de los datos mediante la Automatización del sistema de Información aplicando machine learning en el área logística en VILLA CHICKEN S.A.C., en donde no contaba con una gestión correcta de los productos perecibles y cárnicos, lo que conlleva a desechar mercadería por tiempo de vida de la misma, generando un desbalance en las compras y ventas.

El objetivo general fue determinar la influencia de la Automatización del sistema de Información aplicando machine learning en el área logística en VILLA CHICKEN S.A.C y de esa manera visualizar el impacto que genera de la misma, gracias a la mejora en la predicción de datos desde el sistema de información.

Como resultado se obtuvo que con la Automatización del sistema de Información aplicando machine learning en el área logística en VILLA CHICKEN S.A.C el puntaje promedio del indicador “Rotación de mercancías” se incrementó en un 2.67%, además y el puntaje promedio indicador “Volumen de compras” se incrementó en un 3.7% y el puntaje del promedio indicador “Crecimiento de ventas” se incrementó en un 14.21%. Por lo tanto, se concluyó que la automatización del sistema de Información aplicando machine learning produce efectos significativos en el área logística de VILLA CHICKEN S.A.C.

En la presente investigación se usó el motor de base de datos SQL SERVER 2014 con el lenguaje de programación PHP para el desarrollo del sistema de información, en cuanto la metodología de desarrollo se utilizó **CRISP-DM**. Se La técnica de predicción utilizada fue Regresión lineal de Machine Learning.

Palabras clave: Sistema de Información, rotación de productos, volumen de compras, aprendizaje máquina, CRISP-DM.

Abstract

This thesis shows the understanding of the business, preparation, modeling and evaluation of data through the Automation of the information system by applying machine learning in the logistics area in VILLA CHICKEN SAC, where it did not have a correct management of perishable and meat products, which entails discarding merchandise for its life time, generating an imbalance in purchases and sales.

The general objective was to determine the influence of the Automation of the Information system by applying machine learning in the logistics area in VILLA CHICKEN SAC and in that way visualize the impact it generates, thanks to the improvement in the prediction of data from the system of information.

As a result it was obtained that with the automation of the information system applying machine learning in the logistics area in VILLA CHICKEN SAC the average score of the indicator "Rotation of goods" was increased by 2.67%, in addition to the average indicator score "Volume of purchases "Increased by 3.7% and the average" Sales growth "indicator score increased by 14.21%. Therefore, it was concluded that the automation of the information system applying machine learning has significant effects in the logistics area of VILLA CHICKEN SAC

In the present investigation, the SQL SERVER 2014 database engine was used with the PHP programming language for the development of the information system, as soon as the development methodology was used CRISP-DM. Se The prediction technique used was Linear Regression of Machine Learning.

Keywords: Information System, product rotation, purchasing volume, Machine Learning, CRISP-DM.

I. INTRODUCCIÓN

1.1 Realidad problemática

La definición de logística encierra los distintos procedimientos con el objetivo de que un producto llegue al consumidor (Economía Simple, s.f.). Lo que involucra todo el movimiento de la materia prima, el desarrollo o creación del producto, como el envío de los productos a las manos finales del consumidor. Podemos entonces asegurar que se encuentra relacionada con acciones tales como la distribución, el movimiento de mercadería, almacenamiento. Ahora bien, la logística está alejada de la creación de un bien, hecho que genera que se realice un outsourcing para esta área (Anaya, 2017, p.47)

En base a lo dicho por GS1 Perú, organismo de origen internacional, esta área tiene un gran impacto en la administración de la empresa, calculándose que entre el 50% al 70% de los costos vienen de esta área (Gestión Logística e Inventarios S.A.C., s.f.). Bajo esta definición entendemos la importancia que puede tener esta área para la empresa, dado que bajo ella radica una gran cantidad de los costos. Además, debemos asumir que las mejores prácticas en esta área, pueden ayudar en gran medida a la empresa, bajo una correcta gestión y planes en función a objetivos que vayan directamente relacionados a un plan de empresa.

La empresa Centro Hogar Chiclayo E.I.R.L. dedicada a la comercialización de colchones, la cual se ha visto afectada durante los últimos años en su desabastecimiento de sus stocks en los almacenes por la falta de un control en sus actividades logísticas, en lo que los problemas de falta de insumos, falta de control de los ingresos y salidas en consecuencia pérdidas en las ventas es constante y las compras de los insumos se ve afectada por las compras al por menor que se realizan, afectando al presupuesto. La empresa no cuenta con procedimientos claros, viéndose necesario hacer una mejora en la restructuración del proceso (Rios Burga, 2016)

La sobrevivencia de una empresa de gran tamaño, está ligada a la capacidad de adaptación que tenga, los cambios referidos a tecnología como culturales están presentes constantemente (Comercio, 2017). Este resulta ser una tarea vital para la empresa, existir o crecer a pesar de la competencia. Siendo necesario para tal tarea la toma de decisiones, que conlleve a mejores oportunidades o crecimiento de ella. La inteligencia de la toma de decisiones es transcendental mediante una estrategia establecida fijamente dentro de una organización, ya para llevar a cabo esta actividad con precisión se involucran vario

criterios, análisis de datos y un estudio exhaustivo den medio en la que se rodea. Así mismo la importancia de conocer el entorno de la organización se convierte en un arma potencial para sobrevivir en el mercado actual, entonces la toma de decisiones mediante una herramienta facilita el crecimiento de la empresa (Montalvo, 2016).

Por lo que la toma de decisiones debe basarse en el conocimiento no solo de la empresa, sino también del entorno, bajo esta misma idea debemos hablar de la innovación que debe contar las empresas para poder estar presentes en el mercado. Para el Director de la Escuela de Postgrado de la UPC, Guillermo Quiroga, la supervivencia de las empresas está referida al compromiso de innovación que las empresas con emprendimiento se han impuesto (Comercio, 2017).

Según la cadena de restaurante de Villa Chicken S.A.C. existe una un desbalance en su área logística, en un principio tenemos que el almacén no cuenta con una gestión correcta de los productos, lo que conlleva a desechar mercadería por el tiempo de vida de la misma, como sobrecarga de la misma de productos que se quedan días en el almacén sin salir y que se acumulan dado que se realiza un pedido de la misma cantidad cada cierto tiempo. Por otro lado, tenemos, que las compras de los insumos no están en relación directa a las ventas de la empresa, en ocasiones se compra más insumos o hacen falta productos que se compra por otros distribuidores minorista con otros precios, para responder a la demanda. Lo que afecta al presupuesto de las compras presupuestadas para los insumos.

Las ventas de Villa Chichen S.A.C. también tienen problemas en el cumplimiento de las entregas a tiempo de los productos, tanto en mesa como en delivery ocasionado generalmente por la falta de manejo de stock, lo que afecta al cliente en su sensación de satisfacción. Así como en las ventas que, al no tener un buen manejo predictivo de las cantidades necesarias para la venta, o conocer las necesidades o gustos de sus clientes en temporadas.

1.2 Trabajos previos

1.2.1 Nacionales

Tesis de Montalvo (2016), titulada “Análisis Comparativo de Técnicas de minería de datos para la predicción de ventas”, teniendo como objetivo evaluar las técnicas de minería de datos Regresión, Series temporales y Redes Neuronales para predecir las ventas de la empresa EL ASTRO SAC. Investigación de tipo aplicada, con un diseño

cuasi experimental. Población tomada fue el total registros de ventas del periodo 2012-2015 inventarios con una muestra del 100% representado con 46122 registros de ventas. Como resultado se arrojó los grados de precisión por cada técnica de predicción, HoltWinters un 84.42%, Holt con un 86.96% Y ETS con 81.45%. En conclusión, Holt es el algoritmo de regresión lineal simple con más grado de precisión para predecir las ventas.

Tesis de Cardeñoso Rivas & Misle De La Torre (2016), titulada “Propuesta de Desarrollo de Pronósticos y Control de Inventarios para la Mejora de la Gestión de Pedidos y Distribución en la Empresa MARLO E.I.R. L, Cusco, 2016”, tuvo como objetivo desarrollar los pronósticos que mejore la gestión de inventarios y pedidos de la empresa MARLO E.I.R.L. El tipo de investigación aplicada, diseño cuasi experimental, La población fue de 430 registros transaccionales del año 2014-2015. Obteniendo como muestra 87 registros. Como resultado después de la implementación, los tiempos de entrega de los pedidos disminuyeron en unos 13.02%, los volúmenes de compras aumentaron en 9.23%, finalmente las ventas con 19.92%. En conclusión, después de la implementación del sistema, los pronósticos de rotación de mercancías disminuyó la pérdida de inventarios a raíz del vencimiento, por consecuencia se optimizó los costos logísticos y tiempos.

Tesis de Mendoza & Anchiraico (2018), titulada “Determinación de Patrones de Ventas utilizando machine learning en boticas independientes para mejorar las Ventas”, teniendo como objetivo predecir la rotación de productos en un determinado tiempo para reducir los vencimientos, Aumentar la rotación de productos y los volúmenes ventas. Se comparó la metodología de minería de datos KDD, SEMMA Y CRISP-DM, por lo cual la metodología escogida fue CRISP-DM. El tipo de investigación aplicada, diseño cuasi experimental, La población fue la cantidad de registros transaccionales de la empresa Botica Amberfama E.I.R.L, representando un total de 465240, conformado por las ventas del año 2017. Como resultado después de la implementación, las ventas aumentaron un 15.45%. En conclusión, La metodología utilizada para el proyecto de minería de datos fue CRISP-DM, así mismo se utilizó la técnica de regresión lineal para predecir los movimientos de productos.

Tesis de Zapata (2017), titulada “Mejora de un sistema de gestión logística para la reducción de los costos en la empresa EYSM INGENIERIA SAC de Callao”, teniendo

como objetivo reducir los costos y optimizar los costos logísticos de la empresa. Investigación de tipo aplicada, diseño cuasi experimental. Población utilizada fue un total de 234 registros de inventarios con una muestra del 100% representado en un periodo de 7 semanas. Como resultado se incrementó la calidad de pedidos generados en un 89.80%. En conclusión, los costos logísticos sin un sistema de información automatizado fueron de 71.68%, Mientras que después de la automatización del sistema los costos logísticos se redujeron en un 13%.

Tesis de Vallejos (2017), con el título “Sistema Web para el Proceso Logístico en la organización Soluciones Dinetech S.A.C”, teniendo el objetivo de dicha investigación se basó en determinar la influencia de un sistema de gestión logística de la organización, el criterio de demostración fue percibir la exactitud de la rotación de inventario. Investigación con diseño pre-experimental. La muestra obtenida fue de 20 registros de inventarios con respecto al exactitud de inventario y referencia de pedido para medir la claridad de pedidos. Como resultado respecto al indicador percibir del inventario se obtuvo un incremento de un 28.48% de un 67.75% a 96.2%, mientras el indicador referencia de los pedidos generados se incrementó un 60%. Se concluyó que el sistema agilizo el proceso logístico en la organización.

Tesis de Ríos (2016), con el título “Propuesta de procedimiento de Gestión Logística en la Empresa Centro Hogar Chiclayo E.I.R.L. para disminuir pérdidas de ventas por roturas de Stock”, teniendo como objetivo mejorar los procedimientos de gestión logística. Empleando para el estudio de información de setiembre del 2014 a agosto del 2015. Consiguiéndose mediante esta tesis, definir procesos de logística, resolver problemas de roturas a través de la regularización de pedidos, control de entradas y salidas, aseguramiento del stock, empleándose un sistema de información.

Tesis Díaz (2017), con el título “Gestión Logística para incrementar los stocks de abastecimiento del área de compras de la empresa Consorcio Linely. Cerro de Pasco, 2017” para la obtención del grado de Ingeniero Industrial, objetivo de la investigación mejorar el balance de stock de abastecimiento de productos del área de compras de Consorcio Linely, Cerro de Pasco, 2017. Investigación de tipo aplicada, cuasi-experimental, teniéndose como población 24 hojas de registros de despacho del almacén. Teniéndose como resultados que si existió una mejora en el stock de abastecimiento de un 0.5668 al anterior de 0.2993. La importancia de las investigaciones realizadas con anterioridad, que están publicadas como diversos formatos de investigación como revistas

científicas o artículos indexados en distintos repositorios de investigación. Para desarrollar un tema de investigación es importante conocer los antecedentes que respalde al tema de estudio.

Tesis de Medrano (2017), con el título “Modelo de minería de datos usando Machine Learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016”. Con el objetivo de dicha investigación fue mejorar el diagnóstico mediante patrones de reconocimiento con modelo de minería de datos. Teniendo como tipo de estudio pre-experimental porque presenta un pre-test y post-test mediante el modelo minería aplicado al reconocimiento. En esta investigación se desarrolló mediante la metodología CRIPS y Machine Learning con el fin posteriores predicciones. En conclusión, logro mejorar el diagnóstico de enfermedades a partido de un modelo creado con Machine Learning de

En la tesis de Andrade (2018), con el título “Inteligencia de negocio del proceso de ventas en la Empresa ENFOCATEC S.A.” con el motivo de obtener el grado académico de magister en Ingeniería de Sistemas de Tecnología de Información en la Universidad Cesar Vallejo; determinó la inteligencia en el proceso de las ventas en la Empresa *ENFOCATEC S.A.* La cual realizo una metodología de investigación aplicada de diseño no experimental, con corte transversal es decir obtienen los datos en un solo momento; con el propósito de mejorar ganancias mediante la toma de decisiones en los distintos niveles de producción de la empresa; se realizó el análisis con población de 120 trabajadores de la empresa ENFOCATEC S.A. donde la muestra estuvo comprendida con el total de población. Los resultados obtenidos de dicha investigación fue el siguiente mediante la inteligencia de negocios se obtuvo el que 46.6% de lo considero bueno y por otro lado el 29,2% considero como regular y un 24,2% considera mala. En Conclusión, si la inteligencia se aplica de forma óptima el resultado será aceptable en distintas áreas de la empresa.

Tesis de Torres (2017), con Título “Sistema Móvil para la Inteligencia de Negocios del proceso de ventas en Schroth Corporación Papelera S.A.C” con el motivo de obtener el grado académico de magister en Ingeniería de Sistemas en la Universidad Cesar Vallejo; la cual determinó que mediante un aplicativo móvil e inteligente mejora un proceso de ventas, de las cuales se enfoca en la variable de tomada de decisiones mediante una inteligencia de negocio, y mejora las dimensiones del sistema de información; el diseño

de estudio de dicha investigación esta enfoca en un diseño pre experimental con un enfoque cuantitativo. La muestra fue de 179 colaboradores que fue determinado por muestreo aleatorio simple; la técnica utilizada para dicha investigación fue la encuesta y bajo el cuestionario graduado por Likert para la variable que fue validada bajo la demostración de prueba de confiabilidad se aplicó mediante expertos y alfa de Cronbach. Los resultados evidenciaron que un sistema de información innova y ayuda a la toma de decisiones mediante la inteligencia de negocios dentro del proceso de ventas en SCHROTH Corporación Papelera S.A.C. En conclusión, los expertos también afirman que los sistemas basados en inteligencia de negocios ayudan a la toma de decisiones a las empresas emprendedoras.

Tesis de Urdy y Ceballos (2017), con el título de “La gestión logística y su influencia en la competitividad en las PYMES del sector construcción, equipos y herramientas del distrito de puente piedra” con el objetivo de investigación fue determinar la influencia de un control de gestión logístico en las pymes de las empresas privadas de importadoras de máquinas pesas del distrito norte de Lima. El tipo de enfoque de estudio de dicha investigación enfoque cuantitativo con diseño experimental-transversal, la población utilizada para dicha investigación de 15 empresas utilizando la muestra de la 100% de la misma. Teniendo como resultado evidente que disminuyó el porcentaje de errores cometidas por dichas empresas.

Tesis de Ávila (2017), con el título “La gestión logística y su influencia en el valor ganado en los proyectos de edificación en la selva peruana de una empresa constructora de Lima Metropolitana”. El objetivo que se planteó fue determinar cuál es la influencia de la gestión logística en el valor ganado de diversos proyectos desarrollados en las empresas constructoras de Lima Metropolitana. Se realizó un estudio cuantitativo que fue efectiva para determinar el valor ganado. Se trabajó con una muestra de 14 colaboradores y 3 directivos. Los resultados obtenidos fueron favorables ya que se logró identificar diversos procesos que influyen directamente con la información y medios de inventario, pedido y transporte; siendo esto un punto importante para la determinación de dicho valor.

Tesis de Parodi (2016), con el título “Gestión administrativa y la gestión logística del hospital de Chancay, año 2016”. El objetivo de dicha investigación determinó la gestión administrativa con respecto a la gestión logística en el hospital Chancay, con un enfoque

cuantitativo, la investigación se realizó con 120 trabajadores de área administrativas, obteniendo una muestra aleatoria simple de 101 trabajadores. Los resultados obtenidos de dicha investigación fue que si existe una relación moderada en las dos variables que es gestión logística y gestión administrativa de un ($r=0.539$).

1.2.2 Internacionales

En el artículo Científico de Cheriyan, Ibrahim, Mohanan y Treesa (2018), con el título “Predicción inteligente de ventas usando técnicas de aprendizaje automático”, El objetivo de dicha investigación fue comparar 3 modelos predictivos para mejorar las predicciones de ventas futuras. Se analizó detalladamente el concepto de datos y previsión en ventas. Se Tomó como muestra ficha de observación considerando ventas trimestrales de los años 2015, 2016 y 2017. Como resultado fue el estudio comparativo entre los 3 algoritmos enfocados en rendimiento de predicción de ventas, es el algoritmo modelo lineal se encuentra mostrando un 98% de precisión general y el segundo algoritmo de aumento gradiente fue de 78%, mientras que el de árbol de decisión con 68% de precisión general, Finalmente se pudo comparar en base a la evaluación empírica los 3 algoritmos el mejor modelo es el modelo lineal.

En el artículo Científico de (VARUNA, KAVITA, & RAMYA, 2016) , con el título “Un análisis comparativo sobre regresión lineal y regresión de vectores de soporte.”, El objetivo de dicha investigación fue comparar 2 modelos predictivos que sean más convenientes para pronosticar las ventas, finanzas, movimientos de inventarios de una series de tiempos. Como resultado se obtuvo que para los datos de series de tiempo se analizan como modelo de regresión lineal. Finalmente, se concluye que la técnica de regresión lineal permite pronosticar con mucha más certeza cuando se quiere evaluar el comportamiento conforme vaya avanzando el tiempo.

En el artículo Científico Bing y Yuliang (2018) de con el título “Predicción de la intención de compra del usuario basada en el aprendizaje automático”, El objetivo de dicha investigación fue introducir el método de árbol de decisiones para tratar el problema de la clasificación de intereses de los usuarios y el uso del sistema web local storage, para obtener los datos necesarios. En el experimento se utilizó el software WEKA para construir el árbol de decisión con el fin de garantizar el nivel de eficiencia del algoritmo. Los datos experimentales se dividieron en tres grupos, Cada grupo tenía 5000 datos experimentales. Como resultado se obtuvo que el tiempo promedio del modelado del

algoritmo comparado con el tiempo fue de 0.41 a 0.36, El algoritmo de árbol de decisión ahorro alrededor de 13%. Finalmente, se concluye que al introducir el método de árbol de decisión se usó para tratar el problema de clasificación de intereses del usuario del sistema.

En la tesis de Gnoza y Barberene, para obtener el grado de Ingeniería de Sistemas con el título “Estudio de factibilidad del uso de Machine Learning con múltiples fuentes de datos en el propósito del tiempo”. El objetivo de dicha investigación fue posibilitar el trabajo I+D en las disciplinas de Machine Learning y IoT y Big Data por tener un gran impacto en la comunidad empresarial aplicadas a distintas áreas de una organización. En la siguiente investigación se aplicó técnicas de diseño predictivo para formas a las nuevas investigaciones con respecto a estas impactantes tecnologías. En conclusión, el estudio descriptivo de dicha investigación demuestra que las nuevas tendencias hacen más fiables a los estudios relacionados con respecto Machine Learning.

En el artículo científico de Mantovani, Horváth, Cerri y Vanschoren (2016), con el título “Afinación de parámetros hiperactivos de un algoritmo de inducción de árbol de decisión”, El objetivo de dicha investigación fue estudiar el aprendizaje automático, Usando el algoritmo de árbol de decisión para detallar los modelos comprensibles y niveles de precisión en varios escenarios. Los experimentos que utilizaron fueron 102 conjuntos de datos heterogéneos analizando el efecto de ajuste en los modelos predictivos. Como resultado fue que presento una mejora estadísticamente significativa.

En la Revista Técnica Energía científico de De la Hoz, De la Hoz y Fontalvo (2018), con el título “Metodología basada en Cadenas de Markov para la Predicción de la Demanda y Toma de Decisiones en el corto plazo” el caso de estudio que se realizó en la Empresa Eléctrica Quito: El objetivo de la investigación realizada es que determinó el pronóstico de la demanda en la potencia eléctrica dentro de un periodo corto de tiempo. Dicha investigación se enfocó a comparar perfiles de acuerdo a la carga eléctrica utilizando el modelo oculto de Markov para el pronóstico dentro de un periodo de tiempo limitado. Esta herramienta se aplicó por 30 días, a sistema de EEQ la eléctrica, y concluyo que la dicha prueba acertó en un 86% del comportamiento de la demanda esperada. El resultado de esta investigación es que mediante el sistema implementado CENACE se pueda realizar la toma de decisiones en corto plazo, optimizando los recursos y generadores

existentes. En conclusión, de una u otra manera se realiza un análisis para la predicción y toma de decisiones del caso de estudio. (Sanchez, 2018)

España, Tesis de Iribarren (2016), con título “Modelo predictivo. Machine Learning aplicado al análisis de datos climáticos por una placa Sparkun”, El objetivo de dicha investigación fue crear un modelo óptimo para la predicción precisa sobre posibles retrasos en los vuelos debido a los cambios climáticos. La siguiente investigación describe que mediante un modelo de machine learning que pudo diferenciar las fases de una elaboración predictiva para anticipar los vuelos perdidos durante el tiempo de cambio del estado climático. Después de ejecutar el modelo creado que tiene como elección del algoritmo más preciso, teniendo como resultado los siguientes, fue mostrando resultados de menor a mayor conforme se realiza en entrenamiento. Finalmente, Logro cumplir con el alcance del objetivo, mostrando resultados óptimas mediante el modelo establecido.

Argentina, Artículo Científico Ruso, et. al. (2016), con el título “Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning”, el objetivo de dicha investigación fue desarrollar modelos predictivos mediante señales involucradas en problemáticas existentes y mostrarse de manera automática. El siguiente artículo menciona que mediante detección de patrones claves el algoritmo que se genere mediante inteligencia artificial de machine learning es capaz de predecir qué sucederá durante un periodo de tiempo. El resultado obtenido en esta investigación es señalar que técnica de predicción automática se debe incluir dentro de una problemática, siendo este un panorama holístico dentro de los sistemas de información convencionales.

Ecuador, Tesis de Valle (2014), con el título “El objetivo que planteo fue proponer un modelo de logístico para el manejo de productos de dicha empresa, por las cuales mejoro la productividad de dicha organización. El método de investigación fue analítico inductivo y deductivo con el fin de determinar la necesidad de la empresa, los resultados obtenidos fueron aprobados y implementados dentro de la misma, se ampliaron los espacios que generaba más congestión de población de improductividad convirtiéndolo en un espacio más óptimo y eficaz para la empresa. Se concluye que para mayor productividad en la empresa es el orden y el buen control de productos.

Revista de Martínez, De La Voz y *eat*. (2017), título de investigación “Gestión logística en Pymes del sector de operadores de carga del Departamento del Atlántico”. El objetivo

que se dio a conocer en esta investigación fue analizar la gestión logística en las medianas empresas de un sector específico que es operadores de carga. Se realizó un tipo de estudio cuantitativo y epistémico positivista, descriptiva, diseño no experimental transaccional. Dicha investigación utilizó una población de 25 directivos y 25 colaboradores. El resultado obtenido en dicha investigación fue lograr certificar a los colaboradores por su alto nivel de eficiencia y respeto y desempeño.

1.3 Teorías relacionadas al tema

1.3.1 Variable Independiente: Automatización del Sistema de Información utilizando Machine Learning

Automatización Sistema de Información

La automatización de un sistema de información es realizar tareas de manera óptima de acuerdo al proceso especificado que la empresa realiza o consulta para dar respuesta a un problema específico.

Según Andreu (1994), menciona que un sistema es "conjunto de procesos referidos a la gestión de información con respecto a las necesidades de una organización, obtiene, elabora y distribuye los datos necesarios para procesos operacionales propios de la organización para apoyar a la toma de decisiones para cubrir las demandas funcionales del negocio de acuerdo a las estrategias" (p.40).

Como se mencionó en la cita anterior la importancia de la gestión de la información radica en la eficiencia de grandes cambios en los volúmenes de datos ingresados mediante el proceso de cada área específica de una organización, con el fin de producir información válida para una toma de decisiones.

Machine Learning

Según Matías (2003), Machine Learning es un conjunto de técnicas que hacen parte de la inteligencia artificial, en base de conjunto de algoritmos tienen como finalidad buscar el aprendizaje de diversos conjuntos de datos, Lo novedoso de los algoritmos es el pronóstico de nuevos escenarios basándose en la experiencia obtenida de los diversos conjuntos de datos para su entrenamiento"(p.4).

Como se mencionó en la cita anterior que el aprendizaje automático se realiza mediante algoritmos de inteligencia artificial con la finalidad de aprender, este se basa en el

creación de programas informáticos que cambian de acuerdo a nuevos datos ingresados de las áreas de operaciones de las organizaciones.

Por otra Mitchell (1997) en su libro indico que "Machine Learning consiste en aprendizaje automático, que detalla una variedad de paradigmas, algoritmos, resultados y aplicaciones, Es un campo multidisciplinario, teniendo como base resultados de inteligencia artificial, estimaciones, probabilidades y estadísticas" (p.14).

Según las citas anteriores, esta nueva tecnología ha impactado a múltiples organizaciones mediante distintos paradigmas o algoritmos de aprendizaje, Machine Learning es una tecnología que se utiliza para facilitar el análisis de datos. Mediante esta tecnología, muchas empresas ven al mundo de forma holística.

Árbol de decisión

Para Bing y Yuliang (2018), "Un árbol de decisión es una herramienta de soporte de decisiones que utiliza un gráfico o modelo de decisiones en forma de árboles y sus posibles consecuencias, incluyendo resultados de eventos aleatorias, recursos, costos y utilidades. Es una forma de representar un algoritmo" (p.44).

Regresión Lineal

Para Bing y Yuliang (2018), "son un conjunto de patrones que ajustan una relación de características dependientes, estableciendo una línea arbitraria que he calculada de una distancia de puntos con valores correspondientes, es una recta lineal que consigue acercarse con mayor precisión al mayor número de puntos" (p.48).

1.3.2 Metodología CRISP-DM

La metodología o modelo de procesos más utilizados en la industria de desarrollo de proyectos es CRISP-DM [CRISP-DM- 2000]. Según Gallardo (2009) en su tesis "*Metodología para la Definición de Requisitos en Proyectos de Data Mining (ER-DM)*" en su (p.15) menciona que "CRISP-DM [CRISP-DM- 2000], guía de referencia más ampliamente utilizada en desarrollo de proyectos de minería de datos [...]" por ello estudiaremos sus 6 procesos detallados para dicha investigación.

CRISP-DM está dividido en 4 niveles de abstracción que se encuentra de forma jerárquica.

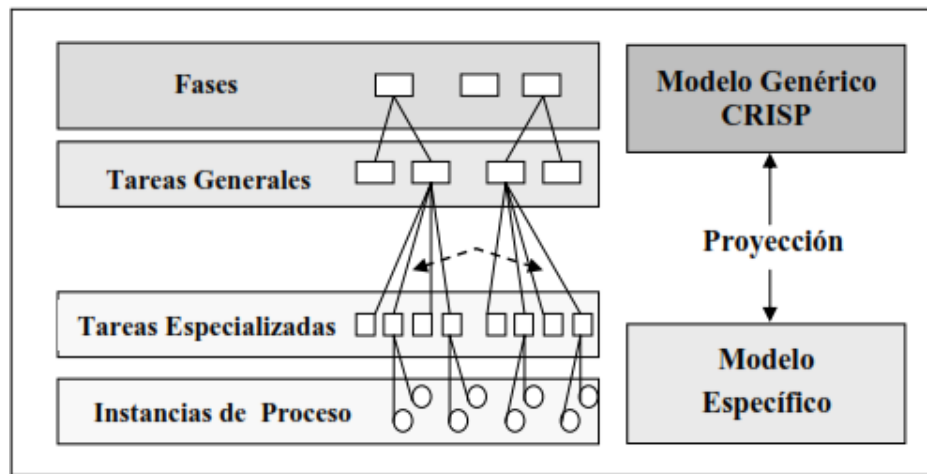


Figura 1 : Esquema de los 4 niveles CRISP-DM

Fuente: (CRISP-DM, 2000). Metodología para la Definición de Requisitos en Proyectos de Data Mining

La Metodología CRISP-DM es una metodología más utilizada por analistas de inteligencia de negocios, para poder desarrollar proyectos de data Warehouse y minería de datos. Según Chapman *et al* (2000) “CRISP-DM consta está organizado en seis procesos: Análisis del Problema, Análisis de Datos, Preparación de Datos, Modelado, Evaluación e Implementación, estas fases se interrelacionan entre sí, además sirve como herramienta para el desarrollo de software y minería de datos” (p.40).

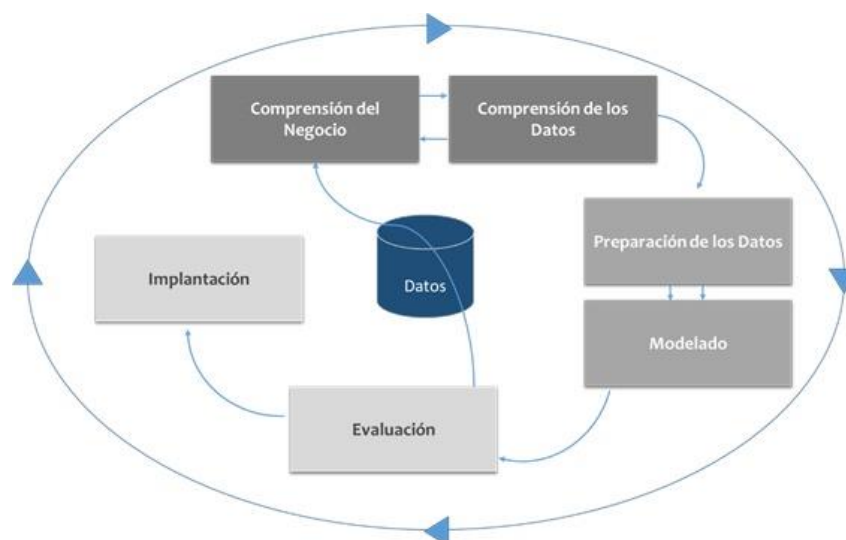


Figura 2: Fases de CRISP-DM

Fuente: Chapman et al (2000). CRISP-DM METHODOLOGY

1. Fase de Compresión de negocio.

El pase de compresión de negocio o reconocimiento del problema, es una de las fases más importantes de la organización. En esta fase se define las metas como necesidades del punto de partida de un proyecto de un punto de perspectiva empresarial o institucional, el fin de esta actividad es de transformar lo convencional en objetivos técnicos o plan de proyecto. Si no se logra entender bien la situación actual del negocio u organización, los algoritmos no serán eficaces, por más que se fueren no se lograra las expectativas del negocio, por ello que es importante reunir a la mayor cantidad de recursos para fomentar la problemática y documentar cada fragmento de la situación para ser analizado y darle un fin la explotación de datos de la organización. En esta fase de debe tener claro cuál es la meta de la empresa, y a dónde se proyecta, gracias a la explotación de los datos históricos. A continuación, se describirá cada actividad de la primera fase de modelo CRISP-DM. (IBM, 2012)

1.1. Determinar los objetivos del negocio

En esta tarea se tiene que tener la máxima información posible de aquellos objetivos comerciales de la minería de datos. Para lograr este punto importante de la organización se define una meta, un problema, criterios y que es lo que se quiere resolver mediante una minería de datos, por ejemplo:

- Predicción de las ventas.
- Rotación de productos
- Crecimiento de las ventas.

En fin, son algunos problemas relacionados a la situación actual de la organización en estudio (IBM, 2012).

1.2. Evaluación de la situación

Se tiene que definir cuál es la situación al principio del proceso de DM, no olvidando los aspectos tales como: ¿Qué datos se tiene para dicho análisis? ¿Hay suficientes recursos para dicho proyecto? ¿Qué factores de riesgo son los más frecuentes o que puede perjudicar al proyecto?, etc. En esta fase se define los requisitos funcionales y no funcionales problema, desde la perspectiva del negocio como de Data Mining (IBM, 2012).

1.3. Determinar los objetivos de DM

Convierte los objetivos del negocio en la meta del proyecto de MD, por ejemplo, consideraremos el objetivo del proyecto en investigación determina la influencia de la metodología CRISP-DM en la toma de decisiones en villa Chicken Jesús María. La meta será determinar la influencia que existe entre la metodología CRISP-DM para la toma de decisiones. (IBM, 2012)

1.4. Realizar el Plan del Proyecto.

En esta tarea ya debe estar listo todo el paso anterior, ya se debe saber con claridad cuál es el objetivo y requisitos de la empresa, considerando lo mencionado se da de alta la primera fase CRISP-DM.

2. Fase de Compresión de Datos

Significa estudiar más de cerca los datos disponibles en dicha organización de minería de datos. Este paso es esencial para evitar problemas de inesperados de inconsistencias para las fases posteriores. En este punto se debe dar el primer contacto con el problema planteado anteriormente, se tiene que familiarizar con la base de datos organizacional y generar una nueva base de datos donde se trabajara la minería de datos. Ya que ha esta se estar realizando consulta con mayor frecuencia. Las tareas relacionadas a esta fase son las siguientes:

2.1 Recolectar los datos iniciales

La primera tarea de la segunda fase es la recolección de datos CRISP-DM, esta tarea tiene con objetivo elaborar un mapo general y realizar una lista de los datos requeridos para dicha elaboración del proyecto de minería de datos, realizando técnicas que deben ser inherentes a la solución del problema.

Así mismo, se debe de determinar los atributos, columnas. Se debe determinar si realmente existen datos consistentes en las tablas existentes.

2.2 Descripción de los datos

Existen diversivas formas de describir los datos, pero la mayor parte de las investigaciones de basa en calidad cantidad de los datos; cantidad de datos disponible para dicho estudio que obviamente tienen que tiene características diversas. Cantidad de datos, la mayor cantidad de modelo de datos tienen equilibrios relacionados, los grandes

volúmenes de datos pueden precisar modelos con mayor eficacia. Tipos de valores, todo lo puede incluir una variable de formatos.

2.3 Exploración de los datos

La exploración de los datos es generar un formato general para todos los datos, esto involucra pruebas de estudios estadísticos, que revelen datos recién procesados, se crean tablas de frecuencia y gráficos de distribución.

2.4 Verificar la calidad de los datos

Esta tarea es la encargada de verificar los datos que se efectúan sobre los datos, la cantidad de valor nulos que se encuentran distribuidos en los campos de una tabla, para encontrar valores que se encuentre fuera de sus especificaciones (IBM, 2012).

3. Preparación de los datos

La preparación de datos es una de los aspectos más importantes de la minería de datos, se estima que el mayor tiempo posible se lleva en esta fase ocupando un porcentaje estimado a 50%-70% del tiempo total. Este proceso se convierte en crítico de acuerdo al objetivo de la organización; las preparaciones de datos pueden implicar las siguientes tareas:

3.1 Selección de datos

Se recolecta un subconjunto de datos, propios de la etapa anterior, donde los criterios de la selección de datos está orientada a los establecidos en la etapa anterior: la calidad, volumen, tipos de datos.

3.2 Limpiar los datos

Esta tarea es la que más tiempo ocupa dentro del proceso, ya que aquí se emplea gran variedad de técnicas con el fin de optimizar la calidad de los datos para la preparación de la fase estructuración de los datos. Entre las técnicas más utilizadas es la normalización y la distinción de tipo de datos, reducción el volumen para su análisis.

3.3 Estructurar los datos

Esta etapa incluye instrucciones de elaboración de los datos como la creación de nuevos atributos de los ya existente, unión de nuevos registros o cambios de valores para atributos ya existentes.

3.4 Integrar los datos

La infracción de datos implica la creación de estructuras nuevas, a partir de los ya elegidos, una demostración seria, el tiempo se puede seleccionar hasta que día o número del año semana o mes etc.

3.5 Formateo de los datos

Esta tarea se encarga de la dar formato a los existentes sin cambiar su estructura inicial en simples palabras transformación sistemática, se puede distinguir por caracteres especiales, puedan utilizar con mayor sencillez y sea factible de utilizar la herramienta (IBM, 2012).

4. Modelo

Para definir el modelo o cuarto proceso de CRISP-DM, simplemente seleccionando las técnicas de modelado más apropiadas para DM y que cumplan con los siguientes criterios:

- Estar acorde al problema
- Contar con datos apropiados y solidos
- Estar a la altura del requisito “cumplir las expectativas del negocio”
- Tiempo apropiado para obtener un modelo
- Conocimiento de técnicas.

4.1 Selecciona la técnica de modelado

Esta tarea consiste en escoger la técnica apropiada para el modelado de DM para resolver el problema afrontado. Es decir, nos enfocamos en el objetivo principal del proyecto y la dependencia con DM existentes. Si el problema es la clasificación, se empleara arboles de decisiones, k-nearest neighbor o razonamiento basado en casos (CBR); sin embargo, si hablamos de predicciones, se tomará el análisis de regresión, redes neuronales; o si el problema es de segmentación, considerar la técnicas de visualización.

4.2 Generar el plan de prueba

Cuando el modelo ya se encuentra construido, se debe elaborar un procedimiento destinado a pruebas de calidad y validez. Por ejemplo, utilizar un margen de error es común ya que es la medida de la calidad DM. Por lo general se dividen los datos en dos conjuntos, uno de entrenamiento y otro de pruebas, los cuales se emplearan para generar modelos.

4.3 Construir el modelo

Una vez seleccionado la técnica, se implementa sobre los datos previamente preparados. Todas las técnicas del modelado poseen un conjunto de parámetros significativos, los cuales definen las características del modelo a generar. La elección de los mejores parámetros es un proceso iterativo y se basa en los resultados generados de las construcciones anteriores. Estos deben ser interpretados y su rendimiento justificado, mediante el análisis de predicción o para la toma de decisiones.

4.4 Evaluar el modelo

En esta fase, los expertos de DM interpretaran los modelos dependiendo del conocimiento existente del dominio y los razones de éxito preestablecidos. Donde, los expertos juzgaran los modelos dentro del contexto del dominio y aplicaran sus propios criterios (seguridad del conjunto de prueba, pérdida o ganancia de tablas (IBM, 2012).

5. Evaluación

Considerado la ejecución de los criterios de éxito de un caso o problema planteado en la primera fase. Además, la integridad calculada se realiza en los datos previamente analizados.

Es necesario evaluar el proceso, asumiendo los resultados en pruebas anteriores, si se presentase alguna observación, será necesario repetir un paso anterior. Además debe tenerse en cuenta que se existen múltiples herramientas para la interpretación de los resultados. Las matrices de confusión son muy usadas en problemas de clasificación y reside en una tabla que indica cuantas clasificaciones se han hecho para cada tipo, llamadas también los checklist, la diagonal de la tabla representa las clasificaciones correctas (IBM, 2012).

5.1 Evaluar los Resultados

En los pasos de evaluados anteriormente, se observaron características tales como la exactitud y generalidad del modelo generado. Esta actividad conlleva a la evaluación del modelo en base a los objetivos del negocio. Evaluándose si es necesario, si el modelo se aleja o es insuficiente, haciéndose pruebas al modelo, en base a un problema.

Además de los resultados directamente relacionados con el objetivo del proyecto, ¿es aconsejable evaluar el modelo en relación a otros objetivos distintos respecto a los originales?, esta información podría ser de mucha importancia

5.2 Revisión del proceso

El proceso de revisión, se describe como la conceptualización del proceso completo de DM, con el fin de reconocer los posibles elementos que pudieran ser renovados.

5.3 Determinar próximos procesos

Si las fases han generado resultados convenientes, entonces se podría seguir a la siguiente fase, sino se sugiere ir a la fase de iteración, fase de preparación de datos. Incluso es posible iniciar un proyecto de DM nuevo.

5.4 Implementación

Cuando el modelo ha sido elaborado y validado por los expertos, se convierte la noción obtenido en ejecuciones dentro del proceso de negocio, ya sea que el analista sugiera otras tareas establecidas en la recolección de datos del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso, como el análisis de riesgo crediticio y detección de fraudes.

Metodología SEMMA

SAS Institute, desarrollador de esta metodología, lo puntualiza con un proceso de adecuado a selección, búsqueda y modelo de grandes volúmenes de datos, con el fin de descubrir patrones de negocios que aún no son conocidos. Esta metodología cuenta con 5 fases básicas del proceso.



Figura 3 : Fases de la metodología SEMMA

Fuente: SAS Institute. Data Mining and SEMMA

El proceso inicia desde un punto clave, que es el corte de una muestra propia de la población, bajo la cual se realizara un análisis. Teniendo como punto de inicio la obtención de la muestra respectiva del problema en estudio. La representatividad es la parte esencial ya que al no cumplirse nula al modelo y los resultados dejan de ser confiables. El tipo de selección de muestreo es al azar ya que cada uno de los individuos tiene la oportunidad de ser parte del muestreo. Esta metodología escoge un muestreo que se le denomina muestreo aleatorio simple.

La metodología SEMMA define que cada muestra obtenida debe tener su propio análisis del proceso y debe ser asociado con el nivel de confianza de la muestra.

La tercera fase de la metodología consiste en la utilización de datos, en base a la exploración realizada en las anteriores fases, en esta fase se define que el dato cumpla con el formato adecuado de análisis que serán ingresados al modelo.

Una vez definido las entradas al modelo, cumpliendo con el formato correspondiente para la aplicación de técnica de modelado, seguidamente se procede con la verificación y análisis y modelado de datos. El objetivo de esta fase es conseguir la relación que existe entre la variable de explicación y variable de objeto de estudio.

La técnica de modelo de datos utiliza la estadística para determinar el respectivo análisis. Finalmente, la última fase consiste en la valoración de los resultados obtenidos mediante el análisis realizado en el modelo, donde tiene sus variables de entrada que son los datos consistentes, se procesa mediante el modelo y finalmente la valoración de la variable de salida.

Metodología KKD

Miguel (2013) indico, que el descubrimiento de conocimiento en bases de datos (KDD), tiene como inicio, la obtención de conocimiento en diversas de bases de datos. Este proceso de obtención de conocimiento, está dividido en diversas etapas, tales como la elaboración de los datos hasta la interpretación de los mismos. Siendo el KDD un proceso

iterativo e interactivo. El primer caso, dado que la salida de alguna de las fases puede descender a otras fases o pasos, los cuales resultaran algo habitual, si el fin es obtener conocimiento de calidad. Mientras que, serán interactivo a los expertos del dominio del problema, porque deben los datos apoyar a la validación del conocimiento obtenido.

El modelo de proceso KDD se simplifica en las siguientes cinco fases resumidas:



Figura 4: Fases de la metodología KDD

Fuente: Moine (2013). Metodologías para el descubrimiento de conocimiento en bases de datos:
un estudio comparativo

Debidamente el modelo define que la minería de datos es la fase dentro del proceso en la cual se realiza la extracción de patrones a partir de los datos. No Obstante, En la actualidad, la comunidad científica, el término KDD y minería de datos se utilizan progresivamente para hacer relacionada al proceso completo de descubrimiento de conocimiento.

El proceso KDD puede simplificarse en las cinco fases mencionadas anteriormente, Fayyad establece nueve etapas para llevarlo a cabo:

1. Comprensión del dominio de aplicación.

En esta primera etapa se recopila toda la información y conocimiento disponible sobre el dominio del sistema, asimismo determinar los fines del proceso KDD desde el enfoque del usuario.

2. Creación del conjunto de datos.

Esta etapa consiste en elegir los datos que se utilizaran para la integración, asimismo para la selección que conformaran la vista de minería de datos. Este paso es necesario para la elaboración de una fuente de datos.

3. Limpieza y pre-procesamiento de los datos.

Esta etapa se lleva a cabo actividades como depuración, depuración de datos anómalos y encontrar los datos faltantes.

4. Reducción y proyección de los datos.

En este paso se identifican atributos importantes para la conceptualización de los datos, puede ser descripción o predicción. Además, se recopila métodos de reducción de dimensionalidad y asimismo métodos de transformación de los datos para decrecer la multitud de variables en el debate para localizar los modelos inconsistentes de los datos.

5. Determinar la tarea de minería de datos.

Se definirá la tarea de minería, enfocándose en el estudio (como agrupamiento, regresión, clasificación o asociación) de acuerdo a lo establecido en la definición de alcance en la etapa 1.

6. Determinar el algoritmo de minería.

Una vez elegida la tarea de minería de datos, en esta etapa se desarrolla el algoritmo para la búsqueda de patrones en los datos. Asimismo, se establecen los modelos y parámetros son mejores dependiendo el grado del problema y los datos que se tienen disponibles.

7. Minería de datos.

Etapa en la que se emplean los algoritmos y técnicas obtenidas al conjunto de datos en búsqueda de los patrones de interés.

8. Interpretación.

Consiste en la explicación de los patrones obtenidos, visualizando y traduciendo el lenguaje técnico a lenguaje entendible por el usuario.

9. Utilización del nuevo conocimiento.

En esta fase se implementa el nuevo conocimiento obtenido, Con esto servirá de apoyo para la toma de decisiones a los directivos. Además, incluye la verificación de potenciales conflictos con conocimientos descubiertos. Finalmente. La metodología KDD muestra las fases generales del proceso de minería de datos, pero no explica que actividades específicas de debe realizar por cada uno, quedando a criterio de los equipos de trabajo.

Metodología CATALYST

Moine y Haedo (2015) encontraron lo siguiente:

La metodología Catalyst, Más conocida como P3TQ (producto, Lugar, Precio, Tiempo y cantidad). Fue implementada por Dorian Pyle en el año 2003 en su libro llamado “Business modelling and data mining”. Dicha metodología Establece la implementación de dos modelos: El claro m modelado del negocio, especifica claramente que y como se tiene que hacer, Explica una serie de pasos para poder detallar los problemas y requerimientos de una organización. El modelo de explotación de información detalla procedimientos determinados a seguir para la elaboración e implementación de modelos de minería de datos en base al modelo de negocio (p.4)

La metodología Catalyst está conformada por 2 partes: Metodología para el modelado del negocio y minería de datos.

Metodología para el modelado del negocio

En esta primera parte se muestran una serie de pasos para el modelamiento del problema y la definición de alcance del negocio. Además incluye diferentes escenarios en el proyecto de minería de datos, favoreciendo una serie de pasos que se lleva a cabo en cada escenario. Pyle (2003) define cinco situaciones o puntos de vistas diferentes para el proyecto:

Escenario 1: Datos

Buscar los datos relevantes que se van a trabajar

1. Determinar el repositorio en donde se recolectarán los datos
2. Identificar los stakeholders (Interesados, directivos, etc.) del proyecto
3. Debatir el Alcance del proyecto con los stakeholders.
4. Caracterizar los conjuntos de datos relacionados al P3TQ
5. Identificar las características de la motivación del negocio para el proceso de recolección de datos.
6. Descubrimiento del problema

Escenario 2: Problema/ Oportunidad

De acuerdo a la realidad problemática del negocio, ver la influencia de la minería de datos para solucionar dicho problema.

1. Identificar a los stakeholders.
2. Describir la situación actual del negocio
3. Identificar los objetivos primordiales con los stakeholders.
4. Definir los datos que se procesaran
5. Modelar el caso de negocio
6. Presentar el modelo de caso de negocio al interesado
7. Describir la situación actual del negocio para el proceso de minería de datos
8. Establecer los requerimientos para implementar.

Escenario 3: Prospección

Descubrir donde la minería de datos puede dar un valor agregado o aporte a una organización

1. Caracterizar las relaciones P3QT importantes de la organización
2. Mapear los flujos principales de la organización
3. Identificar los stakeholders
4. Entrevistar a los stakeholders
5. Identificar los cambios estratégicos que puedan ser de mayor interés para los interesados.
6. Mapear los modelos de minería de datos que pueden dar soporte a los cambios estratégicos de la organización.
7. Gestionar los datos
8. Presentar el caso de uso de negocio a los stakeholders
9. Establecer los requerimientos para implementar

Escenario 4: Modelo definido

Hacer el uso de minería de datos para la construcción de un modelo para una situación específico

1. Identificar los stakeholders
2. Debatir los requerimientos con los stakeholders
3. Definir la situación actual del negocio
4. Identificar los datos para procesar
5. Establecer los requerimientos para implementar

Escenario 5: Estrategia

Dado a una realidad problemática, definir una situación estratégica para analizar si la minería de datos puede ser recomendable con la situación actual del negocio e identificar las opciones para resolver dicho problema.

1. Identificar a los stakeholders.
2. Entrevistar al personal interesado.
3. Identificar la situación actual del negocio
4. Desarrollar mapa de escenarios estratégicos con el interesado.
5. En base al mapa, Crear un modelo de situación estratégica
6. Identificar las características más importantes de las P3TQ
7. Relacionar el mapa de estratégica con las P3TQ
8. Simular errores y descubrir relaciones
9. Identificar las relaciones de P3TQ que puedan ser de mayor interés para el interesado.
10. Describir las fuentes de datos
11. Relacionar cada Problema con el modelo estratégico, asimismo incluyendo los riesgos.

Tomando estos 5 escenarios, El autor Pyle propone una serie de procesos y herramientas para lograr identificar el problema y los requerimientos organizacionales que será el proyecto, asimismo con los datos que se van a analizar.

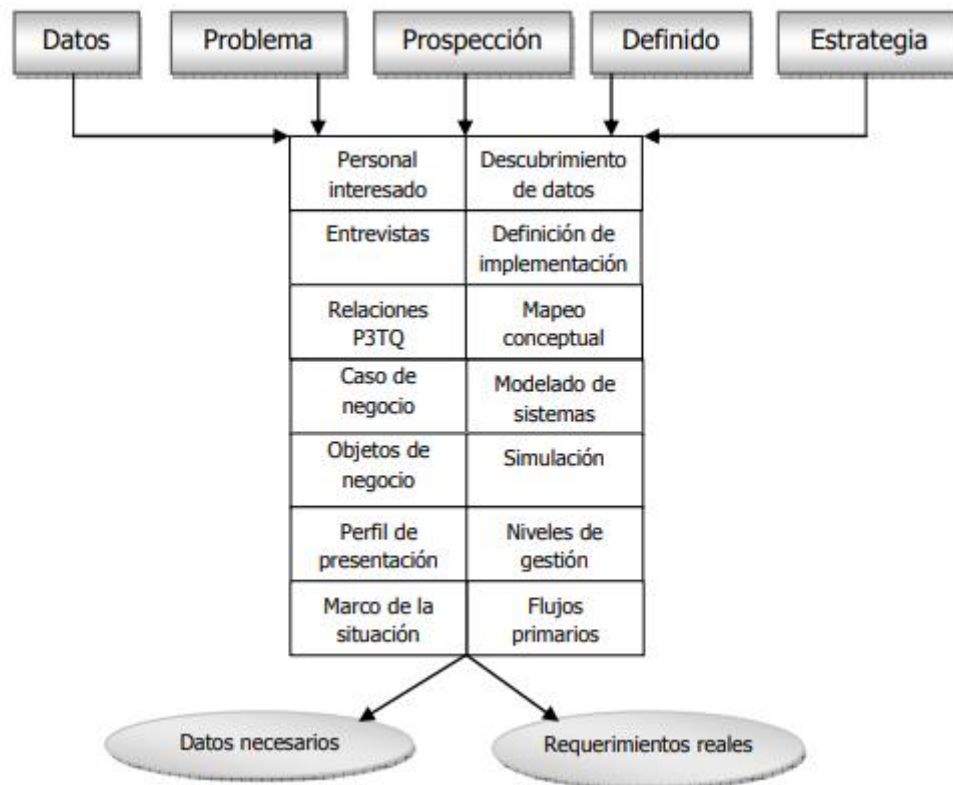


Figura 5: Escenarios de la metodología KD

Fuente: Moine (2013). Metodologías para el descubrimiento de conocimiento en bases de datos:
un estudio comparativo

Tal como muestra la figura, Cualquiera sea el escenario como punto de inicio, Siempre se identificará el problema que el proyecto tendrá obteniendo 2 resultados:

1. Los datos necesarios que se procesaran.
2. Los requerimientos necesarios

Metodología para la minería de datos

Menciona las series de pasos para el descubrimiento de patrones y relaciones de acuerdo a la realidad problemática del negocio.

Preparación de los datos

1. Estimar las variables de estudio (medida de dispersión, datos faltantes, etc).
2. Evaluar los problemas en las variables

3. Chequear variables que no aporten mucha información
4. Identificar que haiga grandes volúmenes de datos
5. Chequear la necesidad de actualizar las variables

Selección de herramientas y modelado inicial

1. Estructurar los datos necesarios para el proceso de minería
2. Identificar las variables de entrada y salida
3. Seleccionar el algoritmo de minería de datos a utilizar
4. Explorar los datos faltantes
5. Crear un modelo inicial (Exploratorio, Clasificación o predicción).

Refinar el modelo

1. Si el método es exploratorio, Definir los resultados encontrados
2. Si el modelo es clasificación o predictivo, verificar la capacidad predictiva del dicho modelo
3. Analizar el modelo con lo stakeholders.

Implementar el modelo

1. Si el modelo es exploratorio, se debe realizar los requerimientos, Luego hacer un informe indicando los resultados obtenidos.
2. Si el modelo es clasificación o predictivo, se deben identificar los requerimientos de implementación realizados antes de la minería de datos.
3. Comunicar a lo stakeholders

Comparación de Metodologías

Tomando en cuenta las etapas generales que componen la minería de datos tales como el análisis, selección, preparación, modelado, evaluación e implementación. Existiendo una relación entre las etapas, de los diferentes enfoques, KDD, CRISP-DM y Catalyst, visto en la figura 6. Por otro lado SEMMA, se orienta a volúmenes mayores de datos.

Fases	KDD	CRISP – DM	SEMMA	CATALYST
Análisis y comprensión del negocio	Comprensión del dominio de aplicación	Comprensión del negocio		Modelado del negocio
Selección y preparación de los datos	Crear el conjunto de datos	Entendimiento de los datos	Muestreo	Preparación de los datos
	Limpieza y pre-procesamiento de los datos	Preparación de los datos	Comprensión	
	Reducción y proyección de los datos		Modificación	
Modelado	Determinar la tarea de minería	Modelado	Modelado	Selección de herramientas y modelado inicial
	Determinar el algoritmo de minería			
	Minería de datos			
Evaluación	Interpretación	Evaluación	Valoración	Refinamiento del modelo
Implementación	Utilización del nuevo conocimiento	Despliegue		Comunicación

Figura 6 Correspondencia entre las fases de cada metodología

Fuente: Moine (2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo

KDD Y SEMMA solo propone actividades generales del proyecto de minería de datos sin explicar de qué forma se deben llevar a cabo cada misma.

Las diferencias de estas metodológicas radican principalmente en la función al contexto en las cuales serán implementas, con diferencias en los requerimientos, los niveles de riesgos que pueda tener cada proyecto de minería de datos. De las 4 metodologías Presentadas anteriormente. Para este proyecto de investigación se utilizará la metodología CRISP-DM yaqué proporciona confiabilidad, robustez y estandarización en los procesos, yaqué especifica detalladamente cada tarea por fase y como se deben ejecutarlas. Asimismo, facilita la creación de futuros proyectos de minería de datos que posean características similares.

1.3.3 Logística

Según Anaya (2007) “La palabra logística, que en transcurso del tiempo proviene del griego (flujo de materiales), se inicia a aplicar en la empresa a partir de la década de los sesenta, [...] la logística se relaciona de una forma directa con todas las actividades

inherentes a los procesos de aprovisionamiento, fabricación, almacenaje y distribución de productos” (p. 47).

La definición de logística entonces abarca tantos los procesos de planificación, implementación, control del flujo, almacenaje, inventarios, así como también la información relacionada desde las compras y ventas.

Inventario

Para (Rojas, 2009) y Goicochea (2009), “El inventario es un bien tangible que debe existir en las empresas para ser consumidas en producción de bienes y servicios, los inventarios comprenden los activos de la empresa de producto en proceso y/o terminado además está comprendida para las ventas, el manejo de inventario ayuda con el control de proceso de la empresa con el fin de entregar un estado contable confiable” (p.39).

Como se mencionó en la cita anterior la importancia del inventario en las empresas, también se puede decir que el inventario contribuye con las partidas del activo corriente de las listas de las ventas, es decir es el valor o precio base del producto o servicio de los almacenes que están en adquisición para dicha elaboración de venta.

Rotación de Inventario

Según Mora (2016) “la rotación de inventario es la Proporción de ventas con existencias promedio. Indica la recuperación de lo invertido en capital través de las ventas” (p.49).

$$\text{Valor} = \frac{\text{Rotacion de productos}}{\text{Total Ventas}} * 100$$

Compras

Según (Anaya Tejero, 2007), “La compra y abastecimiento consiste en un factor clave para la gestión de la cadena de suministros de la organización, donde se pueden controlar y negociar alcanzando alianzas estratégicas” (p.31). Consiste en adquirir un producto o servicio entre consumidores y clientes para realizar la venta.

Volúmenes de compra

Según Mora (2016) “Los volúmenes de compra consiste en el Porcentaje sobre las ventas de los soles o dólares gastados en compras” (p.40). el objetivo es Conocer y controlar el crecimiento de compras y su evolución con la relación a los volúmenes de ventas.

$$\text{Valor} = \frac{\text{VALOR DE COMPRA}}{\text{TOTAL VENTAS}}$$

Ventas

Margen de contribución

Según Mora (2016),” Calculo el porcentaje real de los márgenes de rentabilidad de cada referencia o grupo de productos”

$$\text{Valor} = 1 - \left(\frac{\text{costes variables}}{\text{precio}} \right) * 100$$

1.4 Formulación del problema

Sobre la base de realidad problemática presentada se planteó los siguientes problemas de investigación:

1.4.1 Problema general

¿En qué medida influye la Automatización del sistema de Información aplicando Machine Learning en el área logística en Villa Chicken S.A.C.?

1.4.2 Problemas específicos

Los problemas específicos de la investigación fueron los siguientes:

- ¿En qué medida influye la Automatización del sistema de Información aplicando Machine Learning en el inventario del área logística en Villa Chicken S.A.C.?
- ¿En qué medida influye la Automatización del sistema de Información aplicando Machine Learning en las compras del área logística en Villa Chicken S.A.C.?
- ¿En qué medida influye la Automatización del sistema de Información aplicando Machine Learning en las ventas del área logística en Villa Chicken S.A.C.?

1.5 Justificación del estudio

1.5.1 Justificación teórica

Según Bernal (2010), “La justificación teórica es el propósito de un estudio basado en un conocimiento existente, para poder confrontar una teoría y verificar los resultados”. Asimismo, la justificación teórica pretende generar debate académico en base al conocimiento existente y reflexión. Bajo esta realidad, un sistema de información basado en Machine Learning permitirá conocer si tiene cambios positivos para el área logística o si afecta en manera negativa.

1.5.2 Justificación Práctica

Según Bernal (2010), “La justificación práctica se realiza cuando la elaboración de la investigación permita resolver un problema y proponer estrategias que puedan aplicarse de acuerdo al tipo de estudio que se enfrenta una problemática”.

A partir de lo mencionado, puedo mencionar que el estudio de investigación será un insumo para las siguientes investigaciones.

1.5.3 Justificación metodológica

Según Bernal (2010), “La justificación metodológica del estudio se procede en el proyecto de investigación que se va realizar el planteamiento de un nuevo método o nueva estrategia para generar conocimiento efectivo y veraz”.

Al automatizar el sistema de información aplicando machine learning en el área logística de VILLA CHICKEN S.A.C., se busca por medio de la ciencia y sus métodos, eventos que pueden ser estudiadas por dicha ciencia. Al ser demostrada la viabilidad y fiabilidad de la metodología, se utilizará en otros proyectos similares.

1.5.4 Justificación Social

La puesta en marcha del sistema de información utilizando machine learning en el área logística, servirá como herramienta de gestión para la toma de decisiones con mayor efectividad. Ayudando de esta forma capacitar a otras organizaciones sobre machine learning en el área logística.

1.5.5 Justificación tecnológica

El presente trabajo tiene como finalidad automatizar un sistema de información aplicando machine learning en el área logística de VILLA CHICKEN S.A.C

1.6 Hipótesis

1.6.1 Hipótesis general

HG: La automatización del sistema de Información aplicando Machine Learning produce efectos significativos en el área logística Villa Chicken S.A.C.

1.6.2 Hipótesis específicas

HE1: La automatización del sistema de Información aplicando Machine Learning produce efectos significativos en el inventario del área logística en Villa Chicken S.A.C

HE2: La automatización del sistema de Información aplicando Machine Learning produce efectos significativos en las compras del área logística en Villa Chicken S.A.C

HE3: La automatización del sistema de Información aplicando Machine Learning produce efectos significativos en las ventas del área logística en Villa Chicken S.A.C

1.7 Objetivos

1.7.1 Objetivo general

Determinar la influencia de la Aplicación de sistema de información utilizando Machine Learning en el área logística en Villa Chicken S.A.C.

1.7.2 Objetivos específicos

Los objetivos específicos fueron los siguientes:

OE1: Determinar la influencia de la Automatización del sistema de Información aplicando Machine Learning en el inventario del área logística en Villa Chicken S.A.C.

OE2: Analizar la influencia de la Automatización del sistema de Información aplicando Machine Learning en las compras del área logística en Villa Chicken S.A.C.

OE3: Evaluar la influencia de la Automatización del sistema de Información aplicando Machine Learning en las ventas del área logística en Villa Chicken S.A.C.

II. MÉTODO

2.1 Diseño de la investigación

2.1.1 Enfoque de la investigación

“El enfoque cuantitativo es un proceso secuencial y probatorio. De las diversas etapas que se relacionan entre sí y no se puede adelantar o retrasar pasos, el orden es secuencial, es por ello que se puede redefinir algunas fases. Una vez detallada y delimitada, se elaboran objetivos o preguntas de investigación, revisar literaturas y se elabora un conjunto de procedimientos o una perspectiva teórica “(Hernández, 2014, p.4).

El presente trabajo contiene el enfoque cuantitativo, debido a que obtiene data que será analizada.

2.1.2 Tipo de Estudio

Es una investigación aplicada que consiste en conocer el estado de la sociedad y generar modificaciones para mejorar, causando innovación y actos puntuales en el cambio (Vargas, 2009). La investigación aplicada plantea una mejora de realidad, dando soluciones a los problemas observados. Por lo tanto, podemos afirmar que esta investigación es de tipo aplicada, ya que planteamos una solución a la logística de Villa Chicken S.A.C mediante la aplicación de un proyecto en Machine Learning.

2.1.3 Diseño de la investigación

Se define una investigación que se realiza sin manipular deliberadamente variables. Es decir, no se varía de forma intencional las variables. Lo que se realiza en una investigación no experimental es observar fenómenos que se da en su estado natural; consecutivamente estos cambios son analizados por los expertos en la materia. (Hernández, 2014 p.149)

Según Hernández, Fernández y Baptista (2014) indicaron:

Los diseños preexperimentales también manipulan deliberadamente, al menos, una variable independiente para observar su efecto sobre una o más variables dependientes, sólo que difieren de los experimentos “puros” en el grado de seguridad que pueda tenerse sobre la equivalencia inicial de los grupos. (p. 151).

El diseño de la investigación es pre experimental, debido a que se manipula la variable independiente (Automatización de sistema de información utilizando machine learning) para analizar el impacto que genera en la variable dependiente (Logística).

2.2 Variables, operacionalización

2.2.1 Matriz de operacionalización de las variables

Variables	Definición Operacional	Dimenciones	Indicadores		Instrumento
Dependiente	Según anaya(2007), mencion que "logística se relaciona de una forma directa con todas las actividades inherentes a los procesos de aprovisionamiento, fabricación, almacenaje y distribución de productos"(p.47)	Inventario Rojas (2009,p.39)	Indicador	Formula	Reporte de Registros
Logistica			1. Rotacion de productos (Mora,2016)	Valor= $\frac{Rotacion\ de\ productos}{Total\ ventas} \cdot 100$	
		Compras Anaya(2007,p.24)	1.Volumen de compras (Mora,2016)	Valor= $\frac{Inventario\ Final}{Ventas\ promedio} \cdot 30\ días$ Valor= $\frac{Valor\ de\ las\ compras}{Total\ de\ las\ ventas}$	
		Crecimiento de ventas (Acosta,2011,p.75)	2. Crecimiento de ventas (Acosta,2011)	Valor= $(1 - (\frac{cotes\ variables}{precio})) \cdot 100$	
Variable Independiente					
Automatización de sistema de información aplicando machine learning	Según (Andreu, Ricart, & Valor, 1994), Mencionan que un sistema de información es " un conjunto de procesos orientados a gestión de la información con respecto a las necesidades de una organización, obtiene, elabora y distribuye los datos necesarios para los procesos operacionales de la organización para apoyar a la toma de decisiones para				

Figura 7. Cuadro de Operacionalización de las variables

2.3 Población y muestra

Población

Una población es el conjunto de datos que son obtenidos de un determinado espacio o ambiente de estudio, escogiéndose un número específico para determinar el valor global (Hernandez, 2017)

Según Hernández (2017), la población es definida como un conjunto de elementos relacionados del que se busca tener conocimiento. Este conocimiento, puede manifestarse de diferentes formas estadísticas. Para este proyecto de investigación se utilizó una población de 59 registros del periodo Febrero y marzo del año 2019.

Muestra

Según Hernández (2001), la muestra es una porción de la población, con características semejantes a ella. Asimismo, su tamaño, se centra en métodos estadísticos, con el fin de definir un tamaño en función al universo investigado. Por otro lado, Hernández, Fernández, Baptista (2014) indicaron que la muestra es una parte de la población, y es de ella de la que se obtiene datos.

La muestra estuvo comprendida por el total de población de estudio

$$n = \frac{Z^2 * N * p * q}{e^2 * (N - 1) + (Z^2 * p * q)}$$

Dónde:

N=59 Registros de productos (Población)

Z=Nivel de Confianza (95% = 1.96)

p=Proporción = 0.5

q=Porcentaje que no tiene atributo deseado = 0.5

e= Error máximo aceptable = 0.05

$$n = \frac{(1,96)^2 * 59 * (0,5) * (0,5)}{e^2 * (59 - 1) + (1,96^2 * (0,5) * (0,5))}$$

$$n = 51$$

El resultado de la muestra es de 51 registros, obtenido mediante la fórmula aplicada anteriormente

2.4 Técnicas e instrumentos de recolección de datos, validez y confiabilidad

Registro

La observación es semejante a un registro confiable, admitido, que parte de un acumulado de categorías. (Hernández et. al., 2010)

Ficha de registro

Es el conjunto de registros tomadas de acuerdo a los indicadores que presenta cada objetivo, y posteriormente pasa por la fase de análisis para realizar una retroalimentación o verificar el estado actual con los cambios posteriores que se realizan de acuerdo al modelo de análisis que se presentan en cada indicador propuesto en el caso de estudio.

Validez

Es el nivel de un instrumento que mide la variable que se pretende evaluar, para lo cual se debe tener conocimiento de los rasgos o características se desean estudiar. (Hernandez, 2017). Para validación de este proyecto se utilizará Juicio de expertos.

Confiabilidad

Es un instrumento de medición, que hace referencia al grado en que su reiterativa, es decir que al volver a aplicar sobre los datos, generara los mismos resultados, veraces y coherentes. Un claro ejemplo son los test de inteligencia (Hernández, Fernández y Baptista, 2010).

El registro de observación usado como instrumento, no requiere de los cálculos de confiabilidad, Ya que la información es emitida por el Área logística de VILLA CHICKEN S.A.C

2.5 Métodos de análisis de datos

2.5.1 Estadística Descriptiva

La primera fase es la descripción de los datos, valores o las puntuaciones obtenidas en cada variable seleccionada (Hernandez, 2017).

El método de análisis de dato para nuestra investigación hace referencia a la dimensión ventas, compras e inventario. Para lo cual se usará como instrumento estadístico el programa SPSS Statistics, ya que facilita la gestión de datos ofreciendo formatos de hoja de cálculo simples para el registro de datos, mediante su fácil interacción con el usuario debido a que la interfaz puede ser controlada por un menú de opciones (IBM, 2012).

2.5.2 Estadística inferencial

Va más allá de las descripción datos, se enfoca en la comprobación de las hipótesis y la generación de resultado propios de la muestra seleccionada en función a una población (Hernandez, 2017).

2.6 Aspectos éticos

Esta investigación se rige en todos los criterios del código de ética de la investigación de la Universidad César Vallejo. Así mismo, en función al artículo 14 del código de ética, se obtuvo los permisos necesarios para esta investigación de parte de la institución en la que se desarrolló la presente tesis, colocado en los anexos. Teniéndose fines positivos a esta investigación que pretende mejorar las ventas en función al machine learning.

III. RESULTADOS

3.1 Pruebas de normalidad

La prueba de normalidad que se utilizó fue el método Kolmogorov aplicada a indicador. Ya que la muestra es mayor a 50, Indicaciones:

- Si $n \geq 50$, se emplea el método de kolmogorov- Smirnov
- Si $n < 50$, se emplea el método de shapiro-Wilk.

Entonces considerando las siguientes medidas y obteniendo la prueba de normalidad se puede determinar que:

- Sig < 0.05 , entonces adopta una distribución no normal- Wilcoxon
- Sig ≥ 0.05 , entonces adopta una distribución normal – T student

Donde Sig. es el nivel crítico de contraste. Por lo cual basados en la definición de comprobación de hipótesis tenemos.

3.2 Dimensión 1: Inventario

3.2.1 Indicador 1: Rotación de mercancías

A los puntajes obtenidos en el Pre Test (antes) y Post test (después) se añadió la columna “diferencia”.

	pretest	posttest	diferencia
1	2,82	2,85	,03
2	5,51	7,15	1,64
3	3,77	3,97	,20
4	4,73	4,80	,07
5	2,05	2,94	,89
6	5,88	6,01	,13
7	2,54	2,85	,31
8	2,86	2,94	,08
9	5,45	5,81	,36
10	5,90	6,13	,23
11	7,65	7,80	,15
12	4,79	4,79	,00
13	4,34	5,26	,92
14	3,50	3,64	,14
15	4,40	4,58	,18
16	1,01	1,44	,43
17	5,33	5,35	,02
18	3,34	3,36	,02
19	7,53	7,78	,25
20	1,25	2,22	,97
21	6,45	7,25	,80
22	7,44	7,70	,26
23	4,69	4,69	,00
24	7,81	8,18	,37

Figura 8 Puntajes obtenidos Pre y Test, Post Test

Calculo de Datos Descriptivos

Tabla 1: *Cálculo Estadístico descriptivos - Rotación de mercancías*

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
Pre test	51	1,01	9,90	5,3282	2,10864	4,446
Post test	51	1,44	10,13	5,6047	2,10243	4,420
Diferencia	51	,00	1,64	,2765	,33725	,114
N válido (por lista)	51					

Tabla de Frecuencia

Tabla 2: *Tabla de Frecuencia Pre Test - Rotación de mercancías*

Pre Test				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido 1,01	1	2,0	2,0	2,0
1,25	1	2,0	2,0	3,9
2,05	1	2,0	2,0	5,9
2,54	1	2,0	2,0	7,8
2,82	1	2,0	2,0	9,8
6,47	1	2,0	2,0	70,6
6,82	1	2,0	2,0	72,5
6,91	1	2,0	2,0	74,5
7,02	1	2,0	2,0	76,5
Total	51	100,0	100,0	

Tabla 3 : Tabla de Frecuencias Post Test - Rotación de mercancías

Post Test					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,44	1	2,0	2,0	2,0
	2,22	1	2,0	2,0	3,9
	2,85	2	3,9	3,9	7,8
	7,80	1	2,0	2,0	86,3
	8,07	1	2,0	2,0	88,2
	8,19	1	2,0	2,0	90,2
	8,48	1	2,0	2,0	92,2
	8,99	1	2,0	2,0	94,1
	9,80	1	2,0	2,0	96,1
	Total	51	100,0	100,0	

Histograma

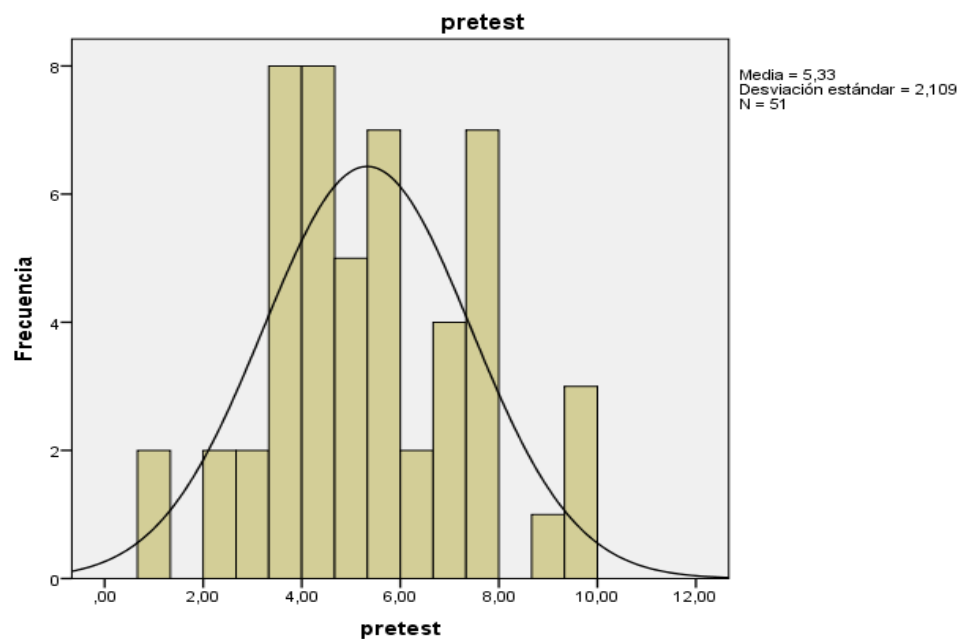


Figura 9 : Gráfico de puntajes obtenidos en el Pre Test- Rotación de mercancías

Como resultado del análisis de Pre-Test en la figura, se representa un histograma de los puntajes obtenidos para el indicador rotación de mercancías, teniendo una media de 5.33, con una desviación estándar de 2.109

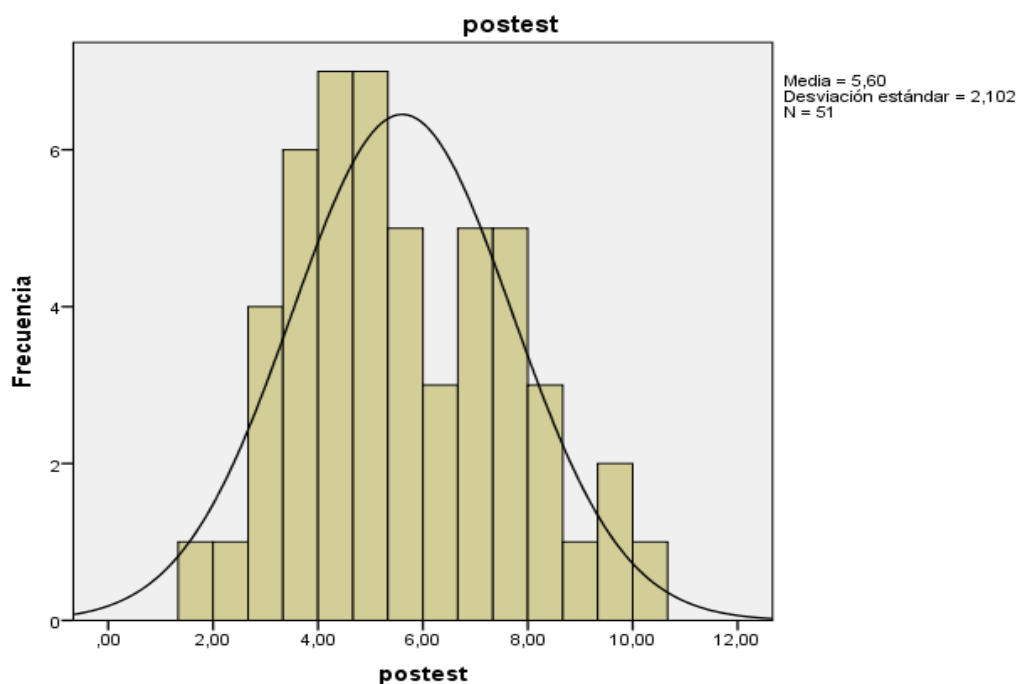


Figura 10: Gráfico de puntajes obtenidos en el Post Test- Rotación de Mercancías

Resultado del análisis de Post-Test en la figura, se representa un histograma de los puntajes obtenidos para el indicador rotación de mercancías, teniendo una media de 5.60, con una desviación estándar de 2.102

Prueba de normalidad

Para determinar si la distribución de la muestra es normal o no, se utilizó la prueba de Kolmorov-Smirnov para una muestra.

Tabla 4: *Prueba de Kolmogorov-Smirnow*

Prueba de Kolmogorov-Smirnov para una muestra

		pretest	posttest	diferencia
N		51	51	51
Parámetros normales ^{a,b}	Media	5,3282	5,6047	,2765
	Desviación estándar	2,10864	2,10243	,33725
Máximas diferencias extremas	Absoluta	,109	,125	,209
	Positivo	,109	,125	,209
	Negativo	-,055	-,066	-,206
Estadístico de prueba		,109	,125	,209
Sig. asintótica (bilateral)		,187 ^c	,045 ^c	,000 ^c

a. La distribución de prueba es normal.

b. Se calcula a partir de datos.

c. Corrección de significación de Lilliefors.

Se observa en la tabla, el valor de significancia (Sig) de la columna diferencia es menos a 0.05, por tal motivo podemos afirmar que el indicador rotación de mercancías sigue en una distribución no normal.

Prueba de hipótesis

Como la distribución de la muestra no es normal, se aplicó una prueba Estadística no paramétrica. La prueba estadística aplicada fue la prueba de Rangos de rangos de Wisconsin. La hipótesis nula y alterna fueron las siguientes:

- **Hipótesis nula (H0):** La automatización del sistema de Información aplicando machine learning **no produce efectos significativos** en el inventario del área logística en VILLA CHICKEN S.A.C
- **Hipótesis Alterna (H1):** La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en el inventario del área logística en VILLA CHICKEN S.A.C

Tabla 5: Prueba de hipótesis – rotación de mercancías

Rangos		N	Rango promedio	Suma de rangos
postest - pretest	Rangos negativos	0 ^a	,00	,00
	Rangos positivos	43 ^b	22,00	946,00
	Empates	8 ^c		
	Total	51		

a. postest < pretest

b. postest > pretest

c. postest = pretest

Tabla 6: Estadística de Prueba Wilcoxon

Estadísticos de prueba ^a	
	postest - pretest
Z	-5,713 ^b
Sig. asintótica (bilateral)	,000

a. Prueba de Wilcoxon de los rangos con signo

b. Se basa en rangos negativos.

Como el valor de Sig. (bilateral) es 0 (menor a 0.05), se rechaza la hipótesis nula y se acepta la hipótesis

Alternativa: La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en el inventario del área logística en VILLA CHICKEN S.A.C

3.3 Dimensión 2: Compras

3.3.1 Indicador 2: Volumen de Compras

Los puntajes obtenidos en el Pre Test (antes) y Post test (después) se añadió la columna “diferencia”.

	PreTest	PostTest	Diferencia
1	25,00	25,44	,44
2	22,88	24,20	1,32
3	34,73	35,92	1,19
4	25,95	25,95	,00
5	28,59	31,05	2,46
6	24,16	24,74	,58
7	23,34	25,83	2,49
8	24,11	24,96	,85
9	24,54	26,72	2,18
10	25,00	25,65	,65
11	24,57	25,22	,65
12	24,64	26,73	2,09
13	25,00	25,28	,28
14	22,64	23,83	1,19
15	24,30	26,96	2,66
16	28,98	31,02	2,04
17	23,50	23,57	,07
18	24,20	24,82	,62
19	20,89	21,10	,21
20	29,67	34,04	4,37
21	25,65	26,93	1,28
22	25,00	26,64	1,64
23	23,48	23,48	,00

Figura 11 Puntajes Obtenidos Pre Test, Post Test y diferencia- Volumen de compras

Cálculo de Datos Descriptivos

Tabla 7: *Cálculos Estadísticos Descriptivos - Volumen de compras*

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
PreTest	51	18,86	37,56	25,8527	3,71210	13,780
PostTest	51	20,31	38,38	27,2727	4,17389	17,421
Diferencia	51	,00	4,92	1,4200	1,11217	1,237
N válido (por lista)	51					

Tabla de Frecuencias

Tabla 8: *Tabla de Frecuencia Pre Test - Volumen de compras*

Pre Test					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	18,86	1	2,0	2,0	2,0
	20,89	1	2,0	2,0	3,9
	21,67	1	2,0	2,0	5,9
	22,47	1	2,0	2,0	7,8
	22,64	1	2,0	2,0	9,8
	22,88	1	2,0	2,0	11,8
	22,94	1	2,0	2,0	13,7
	22,95	1	2,0	2,0	15,7
	23,08	1	2,0	2,0	17,6
	23,14	1	2,0	2,0	19,6
Total		51	100,0	100,0	

Tabla 9 : Tabla de Frecuencia Post Test - Volumen de compras

Post Test					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	20,31	1	2,0	2,0	2,0
	21,10	1	2,0	2,0	3,9
	22,35	1	2,0	2,0	5,9
	23,32	1	2,0	2,0	7,8
	23,48	1	2,0	2,0	9,8
	23,57	1	2,0	2,0	11,8
	23,75	1	2,0	2,0	13,7
	23,83	1	2,0	2,0	15,7
	23,97	1	2,0	2,0	17,6
	24,20	1	2,0	2,0	19,6
Total		51	100,0	100,0	

Histograma

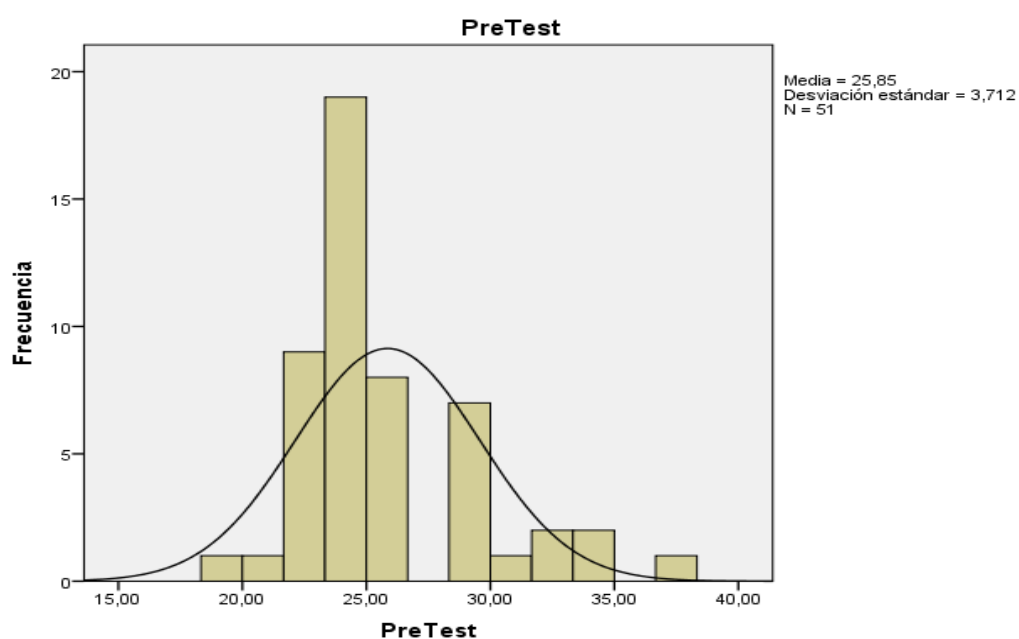


Figura 12 Gráfico de puntuajes obtenidos en el Pre Test - Volumen de compras

Resultado del análisis de Pre-Test que se muestra en la figura, se representa un histograma de los puntuajes obtenidos para el indicador volumen de compras, teniendo una media de 25.85, con una desviación estándar de 3.712.

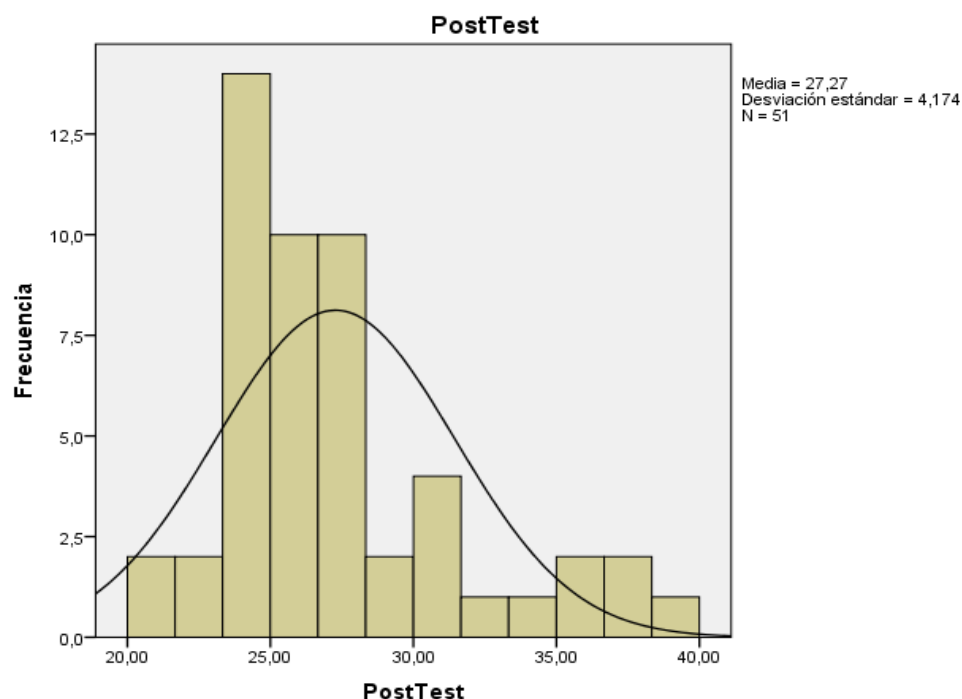


Figura 13: Gráfico de puntajes obtenidos en el Post Test - Volumen de compras

Resultado del análisis de Post-Test se muestra en la figura, se representa un histograma de los puntajes obtenidos para el indicador volumen de compras, teniendo una media de 27.27, con una desviación estándar de 4.174

Prueba de normalidad

Para determinar si la distribución de la muestra es normal o no, se utilizó la prueba de Kolmorov-Smirnov para una muestra.

Tabla 10 : Prueba de Kolmogorov- Volumen de compras

Prueba de Kolmogorov-Smirnov para una muestra				
		PreTest	PostTest	Diferencia
N		51	51	51
Parámetros normales ^{a, b}	Media	25,8527	27,2727	1,4200
	Desviación estándar	3,71210	4,17389	1,11217
Máximas diferencias extremas	Absoluta	,208	,205	,124
	Positivo	,208	,205	,124
	Negativo	-,122	-,113	-,101
Estadístico de prueba		,208	,205	,124
Sig. asintótica (bilateral)		,000 ^c	,000 ^c	,048 ^c

a. La distribución de prueba es normal.

b. Se calcula a partir de datos.

c. Corrección de significación de Lilliefors.

Se observa en la tabla, el valor de significancia (Sig) de la columna diferencia es mayor a 0.05, por tal motivo podemos afirmar que el indicador volumen de compras sigue en una distribución normal.

Prueba de hipótesis

Como la distribución de la muestra es normal, se aplicó una prueba Estadística comparación de medias. La prueba estadística aplicada fue la prueba de muestra relacionadas de T student. La hipótesis nula y alterna fueron las siguientes:

- **Hipótesis nula (H0):** La automatización del sistema de Información aplicando machine learning **no produce efectos significativos** en las compras del área logística en VILLA CHICKEN S.AC
- **Hipótesis Alterna (H1):** La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las compras del área logística en VILLA CHICKEN S.AC

Tabla 11 Aplicación de la muestra relacionadas T student - Volumen de compras

Estadísticas de muestras emparejadas

		Media	N	Desviación estándar	Media de error estándar
Par 1	PreTest	25,8527	51	3,71210	,51980
	PostTest	27,2727	51	4,17389	,58446

Prueba de muestras emparejadas

		Diferencias emparejadas				t	gl	Sig. (bilateral)	
		Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
					Inferior				Superior
Par 1	PreTest- PostTest	-1.42000	1.11217	.15573	-1.73280	-1.10720	-9.118	50	.000

Como el valor de Sig. (bilateral) es 0 (menor a 0.05), se rechaza la hipótesis nula y se acepta la hipótesis alterna: La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las compras del área logística en VILLA CHICKEN S.AC

3.4 Dimensión 3: Ventas

3.4.1 Indicador 3: Crecimiento de Ventas

A los puntajes obtenidos en el Pre Test (antes) y Post test (después) se añadió la columna “diferencia”.

	PreTest	PostTest	Diferencia
1	10,96	18,25	7,29
2	13,98	23,04	9,06
3	11,88	25,20	13,32
4	5,21	20,01	14,80
5	18,39	28,24	9,85
6	9,25	26,52	17,27
7	15,27	27,62	12,35
8	10,47	19,12	8,65
9	23,67	25,47	1,80
10	29,08	43,09	14,01
11	17,62	28,48	10,86
12	27,46	27,46	,00
13	7,38	24,05	16,67
14	29,50	40,12	10,62
15	17,67	25,41	7,74
16	31,35	37,89	6,54
17	25,09	33,92	8,83
18	31,59	46,65	15,06
19	40,41	55,00	14,59
20	30,49	44,51	14,02
21	48,67	56,00	7,33
22	15,98	25,06	9,08

Figura 14: Puntajes obtenidos en el Pre Test, Post Test y la diferencia - Crecimiento de ventas

Cálculo de datos descriptivos

Tabla 12: Cálculo de Datos descriptivos - Crecimiento de ventas

Estadísticos descriptivos						
	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
PreTest	51	5,02	48,67	18,2053	10,00091	100,018
PostTest	51	8,20	56,00	26,1637	11,29393	127,553
Diferencia	51	,00	17,27	7,9584	4,83286	23,357
N válido (por lista)	51					

Tabla de Frecuencia

Tabla 13: *Tabla de Frecuencia Pre Test - Crecimiento de ventas*

Pre Test					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	5,02	1	2,0	2,0	2,0
	5,21	1	2,0	2,0	3,9
	7,25	1	2,0	2,0	5,9
	7,38	1	2,0	2,0	7,8
	8,04	1	2,0	2,0	9,8
	8,73	1	2,0	2,0	11,8
	8,77	1	2,0	2,0	13,7
	9,25	1	2,0	2,0	15,7
	9,51	1	2,0	2,0	17,6
	9,55	1	2,0	2,0	19,6
Total		51	100,0	100,0	

Tabla 14: *Tabla de Frecuencia Post Test - Crecimiento de ventas*

Post Test					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	8,20	1	2,0	2,0	2,0
	9,85	1	2,0	2,0	3,9
	10,48	1	2,0	2,0	5,9
	11,49	1	2,0	2,0	7,8
	11,52	1	2,0	2,0	9,8
	13,79	1	2,0	2,0	11,8
	14,90	1	2,0	2,0	13,7
	15,02	1	2,0	2,0	15,7
	16,36	1	2,0	2,0	17,6
	16,44	1	2,0	2,0	19,6
	Total	51	100,0	100,0	

Histograma

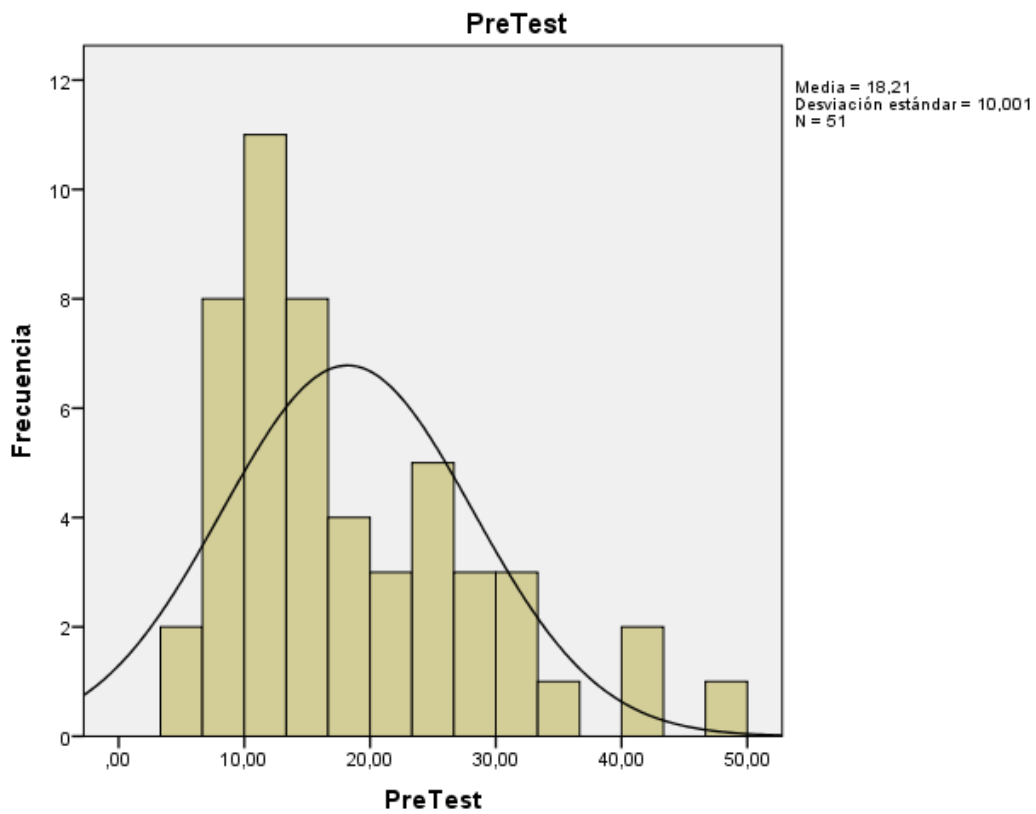


Figura 15: Gráfico de puntajes obtenidos en el Pre Test – Crecimiento de Ventas

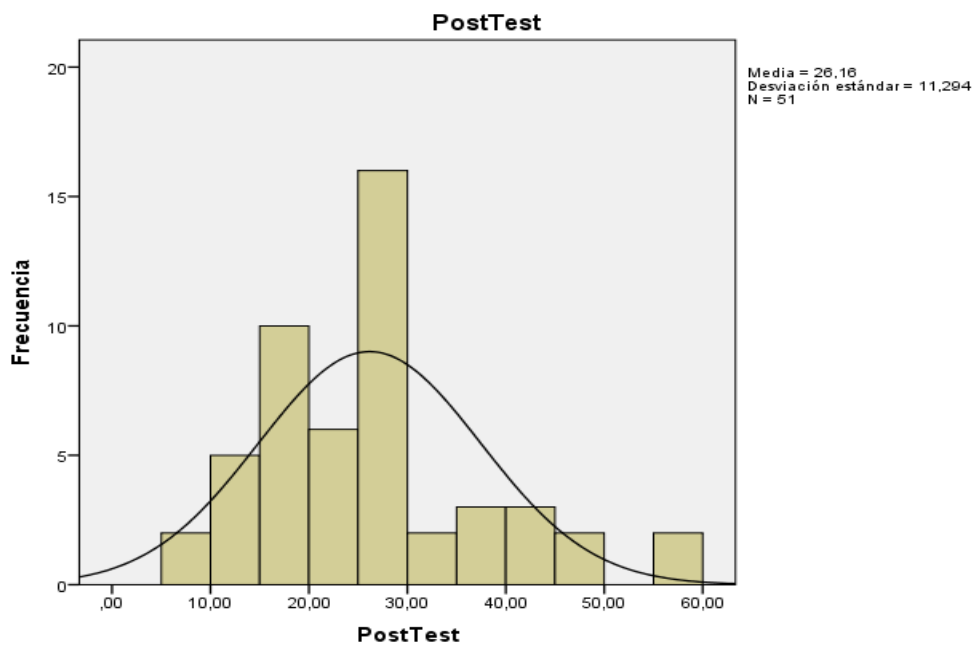


Figura 16: Gráfico de puntajes obtenidos en el Post Test - Crecimiento de ventas

Resultado del análisis de Pre-Test que se muestra en la figura, se representa un histograma de los puntajes obtenidos para el indicador crecimiento de ventas, teniendo una media de 18.21, con una desviación estándar de 10.001

Resultado del análisis que se evidencia de Post-Test en la figura, se representa un histograma de los puntajes obtenidos para el indicador crecimiento de ventas, teniendo una media de 26.16, con una desviación estándar de 11.294

Prueba de normalidad

Para determinar si la distribución de la muestra es normal o no, se utilizó la prueba de Kolmorov-Smirnov para una muestra.

Tabla 15: Prueba de Kolmogorov - Crecimiento de ventas

Prueba de Kolmogorov-Smirnov para una muestra		PreTest	PostTest	Diferencia
N		51	51	51
Parámetros normales ^{a,b}	Media	18,2053	26,1637	7,9584
	Desviación estándar	10,00091	11,29393	4,83286
Máximas diferencias extremas	Absoluta	,149	,133	,095
	Positivo	,149	,133	,075
	Negativo	-,097	-,056	-,095
Estadístico de prueba		,149	,133	,095
Sig. asintótica (bilateral)		,007 ^c	,025 ^c	,200 ^{c,d}

a. La distribución de prueba es normal.

b. Se calcula a partir de datos.

c. Corrección de significación de Lilliefors.

d. Esto es un límite inferior de la significación verdadera.

Como se puede observar en la tabla, el valor de significancia (Sig) de la columna diferencia es mayor a 0.05, por tal motivo podemos afirmar que el indicador crecimiento de ventas sigue en una distribución normal.

Prueba de hipótesis

Como la distribución de la muestra es normal, se aplicó una prueba Estadística comparación de medias. La prueba estadística aplicada fue la prueba de muestra relacionadas de T student. La hipótesis nula y alterna fueron las siguientes:

- **Hipótesis nula (H0):** La automatización del sistema de Información aplicando machine learning **no produce efectos significativos** en las ventas del área logística en VILLA CHICKEN S.AC

- **Hipótesis Alterna (H1):** La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las ventas del área logística en VILLA CHICKEN S.AC

Tabla 16: Aplicación de la muestra relacionadas *T student* - Crecimiento de Ventas

Estadísticas de muestras emparejadas

		Media	N	Desviación estándar	Media de error estándar
Par 1	PreTest	18,2053	51	10,00091	1,40041
	PostTest	26,1637	51	11,29393	1,58147

Prueba de muestras emparejadas

		Diferencias emparejadas				t	gl	Sig. (bilateral)	
		Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
					Inferior				Superior
Par 1	PreTest- PostTest	-7,95843	4,83286	,67674	-9,31769	-6,59917	-11,760	50	,000

Como el valor de Sig. (bilateral) es 0 (menor a 0.05), se rechaza la hipótesis nula y se acepta la hipótesis alterna: La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las ventas del área logística en VILLA CHICKEN S.AC

IV. DISCUSIÓN

En consecuencia, se describen a detalle los resultados obtenidos en la presente investigación al analizar y comparar el comportamiento de la media de los indicadores del inventario, compras y ventas, tales como después de la automatización del sistema de información aplicando machine learning en el área logística en VILLA CHICKEN S.A.C

Partiendo desde la hipótesis específica 1 planteada y luego de evidenciar los cálculos necesarios, se determinó que la media de rotación de mercancías antes de la automatización fue de 5.32 (que representa el 52.61% del puntaje máximo) y luego de la automatización dio como resultado de 5.60 (que representa el 55.28% del puntaje máximo). Entonces partiendo de los resultados se puede afirmar que existe un incremento de 2.67% entre ambos puntajes. Mediante la prueba de hipótesis realizada se rechazó la hipótesis nula, concluyendo que la automatización del Sistema de Información aplicando machine learning **produce efectos significativos** en el inventario del área logística en VILLA CHICKEN S.A.C

En caso de la hipótesis específica 2 planteada y luego de los cálculos necesarios, se encontró que:

En el primer indicador, la media de volumen de compras antes de la automatización fue de 25.85 (que representa el 67.35% del puntaje máximo) y luego de la automatización dio como resultado de 27.27 (que representa el 71.05% del puntaje máximo). A partir de los resultados se puede afirmar que existe un incremento de 3.7% entre ambos puntajes. Con la realización de la prueba de hipótesis se rechazó la hipótesis nula, concluyendo que La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las compras del área logística en VILLA CHICKEN S.A.C

Con respecto a la hipótesis específica 3 planteada y luego de los cálculos necesarios, se encontró que la media de crecimiento de ventas antes de la automatización fue de 18.20 (que representa el 32.50% del puntaje máximo) y luego de la automatización dio como resultado de 26.16 (que representa el 46.71% del puntaje máximo). A partir de los resultados se puede afirmar que existe un incremento de 14.21% entre ambos puntajes. Con la realización de la prueba de hipótesis se rechazó la hipótesis nula, concluyendo que La automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las ventas del área logística en VILLA CHICKEN S.A.C

Estos incrementos de puntajes con respecto al rotación de mercancías, volumen de compras y crecimiento de ventas reflejan el beneficio obtenido por la automatización del sistema de información aplicando machine learning. De la misma forma, Mendoza & Anchiraico (2018) Determinaron los Patrones de Ventas utilizando machine learning en boticas independientes para mejorar las Ventas, Asimismo utilizo la técnica de regresión lineal para predecir las ventas. Los resultados las ventas aumentaron en un 15.45%. Por otro lado, Cardeñoso Rivas & Misle De La Torre (2016) Propusieron un desarrollo de Pronósticos y Control de Inventarios para la Mejora de la Gestión de Pedidos y Distribución en la Empresa MARLO E.I.R. L. Obtenido como resultado el incremento de volúmenes de compra en unos 9.23%, asimismo la rotación de mercancía se incrementó en unos 7.45%.

Por tal motivo se concluye que la automatización del sistema de Información aplicando machine learning produce **efectos significativos** en el área logística de VILLA CHICKEN S.A.C.

V. CONCLUSIONES

- El puntaje promedio obtenido por el indicador rotación de mercancías antes de la automatización fue de 5.32 y luego de la automatización fue de 5.60, que representa un incremento de 2.67% entre ambos puntajes. Con ello se demostró que la automatización del sistema de Información aplicando machine learning **produce efectos significativos** en el inventario del área logística en VILLA CHICKEN S.A.C
- El puntaje promedio obtenido por el indicador volumen de compras antes de la automatización fue de 25.85 y luego de la automatización fue de 27.27, que representa un incremento de 3.7% entre ambos puntajes. Con ello se demostró que la automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las compras del área logística en VILLA CHICKEN S.A.C
- El puntaje promedio obtenido por el indicador crecimiento de ventas antes de la automatización fue de 18.20 y luego de la automatización fue de 26.16, que representa un incremento de 14.21% entre ambos puntajes. Con ello se demostró que la automatización del sistema de Información aplicando machine learning **produce efectos significativos** en las ventas del área logística en VILLA CHICKEN S.A.C
- En consecuencia, demostrando el incremento de puntajes en los indicadores de la dimensión Inventario, **Compras y Ventas**. Se concluye que la automatización del sistema de Información aplicando machine learning produce **efectos significativos** en el área logística de VILLA CHICKEN S.A.C.

VI. RECOMENDACIONES

- Se recomienda definir estrategias de predicción teniendo como base el resultado obtenido con el análisis de predicción de datos a mayor escala, realizar ofertas de productos relacionados por sucursal y estrategias de ventas relacionadas al ofrecimiento de productos adicionales en base al comportamiento determinado en cuenta el algoritmo regresión lineal simple.
- Se recomienda tener en cuenta al ticket promedio como un indicador del consumo de los clientes como capacidad de compra, y no utilizar al ticket promedio como indicador de rentabilidad.
- Se recomienda utilizar predicción de datos para la elaboración de promociones y ofertas de productos. El conocimiento que se obtiene del pronóstico se relacionar productos que los clientes compran en conjunto.
- Se recomienda ampliar esta investigación teniendo en cuenta la problemática de otras organizaciones similares, para que sirva como base de conocimiento para futuras investigaciones

VII. REFERENCIAS

- Anaya, J. (2007). *Logística Integral. La gestión operativa de la empresa*. (3a ed.). Madrid, España: ESIC
- Andrade, M. (2018). *Inteligencia de negocios del proceso de ventas en la Empresa ENFOCATEC S.A., 2018*. (Tesis de posgrado, Universidad Cesar Vallejo, Lima, Perú). Recuperado de http://repositorio.ucv.edu.pe/bitstream/handle/UCV/17683/Andrade_PMY.pdf?sequence=1&isAllowed=y
- Andreu, R. (1994). *Estrategia y Sistemas de Información*. Madrid, España: Mc-Graw Hill
- Bernal, C. (2010). *Metodología de la Investigación*. (3a ed.). Bogotá, Colombia: Pearson Educación
- Bing, L. y Yuliang, S. (noviembre, 2016). *Predicción de la intención de compra del usuario basada en el aprendizaje automático*. Trabajo presentado en el 2016 3ª Conferencia internacional sobre informática suave e inteligencia de la máquina (ISCFI), Dubai. Resumen recuperado de <https://ieeexplore.ieee.org/document/8057447>
- Cheriyian, S. y Ibrahim, S. (agosto, 2018). *Predicción inteligente de ventas usando técnicas de aprendizaje automático*. Trabajo presentado en el 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Reino Unido. Resumen recuperado de <https://ieeexplore.ieee.org/document/8659115/authors>
- De la Hoz, E, De la Hoz, E. y Fontalvo, T. (2019). *Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación*. *Revista Técnica Energía*, 30(1), 247-254. Recuperado de <https://scielo.conicyt.cl/pdf/infotec/v30n1/0718-0764-infotec-30-01-247.pdf>
- Díaz, L. (2017). *Gestión logística para incrementar los stocks de abastecimiento del área de compras de la empresa consorcio Linely* (Tesis de pregrado, Universidad Cesar Vallejo, Lima, Perú). Recuperado de http://repositorio.ucv.edu.pe/bitstream/handle/UCV/1466/D%C3%ADaz_OLM.pdf?sequence=1&isAllowed=y

Economía Simple (s.f.). *Definición de Logística*. Publicado el s. f. Recuperado de <https://www.economiasimple.net/glosario/logistica>

El Comercio. (06 de marzo del 2017). *5 grandes empresas que quebraron por dejar de ser innovadoras*. El Comercio. Recuperado de <https://elcomercio.pe/especial/zona-ejecutiva/actualidad/5-grandes-empresas-que-quebraron-dejar-innovadoras-noticia-1973092>

Gestión Logística e Inventarios S.A.C. (2014). *Gestión Logística e Inventarios S.A.C.* Publicado el s. f.. Recuperado de <http://www.gestionlogistica.com.pe/novedades/importancia-del-area-logistica-en-la-empresa/>

Gnoza, N. y Barberene, M. (2018). *Estudio de factibilidad del uso de Machine Learning con múltiples fuentes de datos en el pronóstico del tiempo* (Tesis de pregrado, Universidad ORT Uruguay, Montevideo, Uruguay). Recuperado de <https://dspace.ort.edu.uy/bitstream/handle/20.500.11968/3761/Material%20completo.pdf?sequence=-1&isAllowed=y>

Gohari, E.Menhri, S. (mayo, 2014). *Un estudio sobre la predicción de la tendencia del usuario hacia las compras en sitios web basados en modelos de comportamiento*. Trabajo presentado en el 2014 6ª Conferencia sobre Tecnologías de la Información y el Conocimiento (IKT), Irán. Resumen recuperado de <https://ieeexplore.ieee.org/document/7030334/authors#authors>

Goicochea, M. (2009). *Sistema de control de inventarios del almacén de productos terminados en una empresa metal mecánica*. (Tesis de pregrado, Universidad Ricardo Palma, Lima, Perú). Recuperado de http://repositorio.urp.edu.pe/bitstream/handle/urp/175/goicochea_ma.pdf?sequence=1&isAllowed=y

Iribarren, D. (2016). *Modelo predictivo. Machine Learning aplicado al análisis de datos climáticos capturados por una placa Sparkfun*. (Tesis de pregrado, Universidad Pontificia Comillas, Madrid, España). Recuperado de <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/14322/TFG000958.pdf?sequence=1&isAllowed=y>

- Mantovani, R y Horváth, T. (octubre, 2016). *Afinación de parámetros hiperactivos de un algoritmo de inducción de árbol de decisión*. Trabajo presentado en el 2016 5ta Conferencia Brasileña de Sistemas Inteligentes (BRACIS), Brasil. Resumen recuperado de <https://ieeexplore.ieee.org/document/7839559/authors#authors>
- Matías, J. (noviembre, 2003). *Boosting con Redes Neuronales RBF. Análisis Sesgo-Varianza en un problema de clasificación*. Trabajo presentado en el VI Congreso Galego de Estadística e Investigación de Operación, España. Resumen recuperado de https://sgapeio.es/INFORMEST/VICongreso/artigos/sesion7_06.pdf
- Medrano, S. (2017). *Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016*. (Tesis de pregrado, Universidad Cesar Vallejo, Trujillo, Perú). Recuperado de http://repositorio.ucv.edu.pe/bitstream/handle/UCV/9852/medrano_ps.pdf?sequence=1&isAllowed=y
- Mitchell, T. (1997). *Machine Learning*. Recuperado de <http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf>
- Montalvo, R. (2016). *Conexión ESAN*. Publicado el 19 de abril del 2016. Recuperado de <https://www.esan.edu.pe/conexion/actualidad/2016/04/19/importancia-toma-de-decisiones-inteligente-estrategica-y-global/>
- Mora, L. (2016). Los indicadores claves del desempeño logístico. Bogota
- Ongsulee, P. (noviembre, 2017). *Inteligencia artificial, aprendizaje automático y aprendizaje profundo*. Trabajo presentado en el 2017 XV Conferencia internacional sobre TIC e ingeniería del conocimiento (TIC y KE), Tailandia. Resumen recuperado de <https://ieeexplore.ieee.org/document/8259629>
- Rios, W. (2016). *Propuesta de procedimientos de gestión logística en la Empresa Centro Hogar Chiclayo E.I.R.L. para disminuir pérdidas de ventas por roturas de stock*. (Tesis de pregrado, Universidad Católica Santo Toribio de Mogrovejo, Chiclayo,

Perú). Recuperado de
http://tesis.usat.edu.pe/bitstream/usat/655/1/TL_Rios_Burga_WinieLisset.pdf

Russo, C., Ramòn, H. Alonso, N. Cicerchia, B. y Esnaola, L. (2016). *Tratamiento masivo de datos utilizando técnicas de machine learning*. Red de Universidades con Carreras en Informática (RedUNCI), (), 131-134. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/52838/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

ScrumStudy (2016). *Una guía para el cuerpo de conocimiento de Scrum (Guía SBOK). Una guía completa para la entrega de proyectos utilizando Scrum*. Arizona, USA: SCRUMstudy

Torres, A. (2018). *Sistema Móvil para la Inteligencia de Negocios del proceso de ventas en Schroth Corporación Papelera S.A.C., Lima - 2017* (Tesis de posgrado, Universidad Cesar Vallejo, Lima, Perú). Recuperado de http://repositorio.ucv.edu.pe/bitstream/handle/UCV/17730/Torres_CAF.pdf?sequence=1&isAllowed=y

Vallejos, M. (2017). *Sistema web para el proceso logístico en la empresa Soluciones Dinotech S.A.C* (Tesis de pregrado, Universidad Cesar Vallejo, Lima, Perú). Recuperado de <http://repositorio.ucv.edu.pe/handle/UCV/1824>

Vargas, A. (1995). *Estadística Descriptiva e Inferencial*. Recuperado de <https://books.google.com.pe/books?id=RbaC-wPWqjC&printsec=frontcover&dq=estadistica+descriptiva&hl=es-419&sa=X&ved=0ahUKEwjg463Tl53cAhVKjlkKHWNsAVMQ6AEIPTAE#v=onepage&q=estadistica%20descriptiva&f=false>

Zapata, A. (2017). *Mejora de un Sistema de Gestión Logística para la reducción de los costos en la Empresa EYSM Ingeniería SAC de Callao, 2017* (Tesis de pregrado, Universidad Cesar Vallejo, Lima, Perú). Recuperado de <http://repositorio.ucv.edu.pe/handle/UCV/1992>

Zuo, Y. y Shawkat, A. (diciembre, 2016). *Predicción de compras al consumidor en una tienda de abarrotes con técnicas de aprendizaje automático*. Trabajo presentado

en el 2016 3er Congreso Mundial de Asia y el Pacífico sobre Informática e Ingeniería (APWC en CSE), Fiji. Resumen recuperado de <https://ieeexplore.ieee.org/document/7941935>

ANEXOS

Anexo 1: Matriz de consistencia

Titulo: Automatización del sistema de Información aplicando machine learning en el area logistica en villa chicken s.a.c					
Problema General	Objetivo General	Hipotesis General	Variables	Dimensiones	Indicadores
¿En qué medida influye la Automatización del sistema de Informacion aplicando machine learning en el area logistica en villa chicken s.a.c.?	Determinar la influencia de la Automatización del sistema de Informacion aplicando machine learning en el area logistica en villa chicken s.a.c.	La automatización del sistema de Información aplicando machine learning produce efectos significativos en el área logistica de villa chicken s.a.c	VARIABLE INDEPENDIENTE: Automatización de sistema de información aplicando machine learning	-	-
Específicos	Específicos	Específicos			Indicadores
¿En qué medida influye la Automatización del sistema de Informacion aplicando machine learning en el inventario del area logistica en villa chicken s.a.c?	Determinar la influencia de la Automatización del sistema de Informacion aplicando machine learning en el inventario del area logistica en villa chicken s.a.c	La automatización del sistema de Información aplicando machine learning produce efectos significativos en el inventario del area logistica en villa chicken s.a.c	VARIABLE DEPENDIENTE: Logistica (Anaya,2007,p.24)	Inventario (Rojas,2009,p.39)	Rotación de productos(Mora,2016,p.4) $\text{Valor} = \frac{\text{Rotacion de productos}}{\text{Total ventas}} \cdot 100$
¿En qué medida influye la Automatización del sistema de Informacion aplicando machine learning en las compras del area logistica en villa chicken s.a.c?	Analizar la influencia de la Automatización del sistema de Informacion aplicando machine learning en las compras del area logistica en villa chicken s.a.c	La automatización del sistema de Información aplicando machine learning produce efectos significativos en las compras del area logistica en villa chicken s.a.c		Compras (Anaya,2007,p.24)	Volumen de Compras (Mora,2016,p.11) $\text{Valor} = \frac{\text{Valor de las compras}}{\text{Total de las ventas}}$
¿En qué medida influye la Automatización del sistema de Informacion aplicando machine learning en las ventas del logistica en villa chicken s.a.c?	Evaluar la influencia de la Automatización del sistema de Informacion aplicando machine learning en las ventas del area logistica en villa chicken s.a.c	La automatización del sistema de Información aplicando machine learning produce efectos significativos en las ventas del area logistica en villa chicken s.a.c		Ventas (Anaya,2007,p.24)	Crecimiento de ventas (Acosta,2011,p.75) $\text{Valor} = \frac{V \text{ actuales} - \text{vaño anterior}}{\text{venta año anterior}} \cdot 100$

Figura 17: Matriz de Consistencia - Automatización de sistema de información aplicando machine learning en el área logística de Villa Chicken S.A.C

Anexo 2: Instrumento de recolección de datos

Tabla 17 Ficha de registros- rotación de mercancías (Pre -Test)

N ° DE FICHA DE REGISTRO:	1
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/02/2019-31/03/2019

Vari able	Indicador	Descripción	Técni ca	Me did a	Instrumen to	Formula
Logística	Porcentaje de rotación de mercancías	Movimiento de mercancías	Observación	Porcentaje	Ficha de observación	$VP = \frac{VENTAS}{I.PROMEDIO} * 100$

FECHA	VENTA	INVENTARIO PROMEDIO	INDICADOR
2019-02-01	12550.11	4457.00	2.82
2019-02-02	47145.79	8557.00	5.51
2019-02-03	11302.42	3000.00	3.77
2019-02-04	20981.08	4433.00	4.73
2019-02-05	8367.78	4078.00	2.05
2019-02-06	23686.76	4026.00	5.88
2019-02-07	6039.23	2378.00	2.54
2019-02-08	11186.27	3915.00	2.86
2019-02-09	21543.86	3953.00	5.45
2019-02-10	19659.91	3335.00	5.90
2019-02-11	23029.42	3010.00	7.65
2019-02-12	27847.98	5809.00	4.79
2019-02-13	10733.49	2473.00	4.34
2019-02-14	8472.69	2424.00	3.50
2019-02-15	42627.21	9696.00	4.40
2019-02-16	2514.03	2477.00	1.01
2019-02-17	19942.47	3745.00	5.33
2019-02-18	12496.78	3738.00	3.34
2019-02-19	24323.43	3230.00	7.53
2019-02-20	4279.68	3419.00	1.25
2019-02-21	15334.9	2379.00	6.45
2019-02-22	26819.8	3604.00	7.44
2019-02-23	19778.14	4216.00	4.69

2019-02-24	20702.01	2651.00	7.81
2019-02-25	12401.22	2813.00	4.41
2019-02-26	21081.33	3093.00	6.82
2019-02-27	13841.41	3966.00	3.49
2019-02-28	30674.99	4173.00	7.35
2018-03-01	12453.1	2285.00	5.45
2019-03-02	27616.04	2869.00	9.63
2019-03-03	11394.96	2889.00	3.94
2019-03-04	15620.74	3709.00	4.21
2019-03-05	14613.79	4293.00	3.40
2019-03-06	30017.93	4277.00	7.02
2019-03-07	23943.72	3701.00	6.47
2019-03-08	29258.55	3030.00	9.66
2019-03-09	22098.98	3198.00	6.91
2019-03-10	22945.1	3047.00	7.53
2019-03-11	45627.21	10102.00	4.52
2019-03-12	23688.28	2669.00	8.88
2019-03-13	25084.83	3170.00	7.91
2019-03-14	19116.11	3523.00	5.43
2019-03-15	16215.45	2752.00	5.89
2019-03-16	14803.21	3463.00	4.27
2019-03-17	15779.68	3605.00	4.38
2019-03-18	23468.14	2371.00	9.90
2019-03-19	48495.69	10455.00	4.64
2019-03-20	13576.89	4043.00	3.36
2019-03-21	32444.01	9582.00	3.39
2019-03-22	28078.21	3986.00	7.04

Tabla 18 *Ficha de registros- volumen de compras (Pre Test)*

N ° DE FICHA DE REGISTRO:	2
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/02/2019-31/03/2019

Variable	Indicador	Descripción	Técnica	Medida	Instrumento	Formula
Logística	Porcentaje de volumen de compras	Volumen de compras	Observación	Porcentaje	Ficha de observación	$VP = \frac{(COMPRA/VENTAS) * 100}{}$

FECHA	VENTA	COMPRAS	INDICADOR
2019-02-01	12550.11	3137.53	25.00
2019-02-02	47145.79	10786.45	22.88
2019-02-03	11302.42	3925.6	34.73
2019-02-04	20981.08	5445.27	25.95
2019-02-05	8367.78	2391.95	28.59
2019-02-06	23686.76	5721.69	24.16
2019-02-07	6039.23	1409.81	23.34
2019-02-08	11186.27	2696.57	24.11
2019-02-09	21543.86	5285.97	24.54
2019-02-10	19659.91	4914.98	25.00
2019-02-11	23029.42	5657.36	24.57
2019-02-12	27847.98	6862	24.64
2019-02-13	10733.49	2683.37	25.00
2019-02-14	8472.69	1918.17	22.64
2019-02-15	42627.21	10356.8	24.30
2019-02-16	2514.03	728.51	28.98
2019-02-17	19942.47	4685.62	23.50
2019-02-18	12496.78	3024.2	24.20
2019-02-19	24323.43	5080.86	20.89
2019-02-20	4279.68	1269.92	29.67
2019-02-21	15334.9	3933.72	25.65
2019-02-22	26819.8	6704.95	25.00
2019-02-23	19778.14	4644.54	23.48
2019-02-24	20702.01	7775.5	37.56
2019-02-25	12401.22	3100.31	25.00
2019-02-26	21081.33	5570.33	26.42
2019-02-27	13841.41	3960.35	28.61

2019-02-28	30674.99	9668.75	31.52
2019-03-01	12453.1	4113.27	33.03
2019-03-02	27616.04	7904.01	28.62
2019-03-03	11394.96	2148.74	18.86
2019-03-04	15620.74	3605.19	23.08
2019-03-05	14613.79	3353.45	22.95
2019-03-06	30017.93	6504.48	21.67
2019-03-07	23943.72	5685.93	23.75
2019-03-08	29258.55	7214.64	24.66
2019-03-09	22098.98	5124.74	23.19
2019-03-10	22945.1	5436.28	23.69
2019-03-11	45627.21	11562.64	25.34
2019-03-12	23688.28	6922.07	29.22
2019-03-13	25084.83	7271.21	28.99
2019-03-14	19116.11	4979.03	26.05
2019-03-15	16215.45	4153.86	25.62
2019-03-16	14803.21	4700.8	31.76
2019-03-17	15779.68	3544.92	22.47
2019-03-18	23468.14	5967.04	25.43
2019-03-19	48495.69	11123.92	22.94
2019-03-20	13576.89	3294.22	24.26
2019-03-21	32444.01	11185.28	34.48
2019-03-22	28078.21	7119.55	25.36

Tabla 19 *Ficha de registros- Crecimiento de Ventas (Pre Test)*

N ° DE FICHA DE REGISTRO:	3
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/02/2019-31/03/2019

Variable	Indicador	Descripción	Técnica	Medida	Instrumento	Formula
Logística	Porcentaje de crecimiento de ventas	Crecimiento de ventas	Observación	Porcentaje	Ficha de observación	$VP = \frac{(VENTAS\ 2019 - VENTAS\ 2018)}{VENTAS\ 2018} * 100$

FECHA	VENTAS 2019	VENTAS 2018	INDICADOR
2019-02-01	12550.11	11310.01	10.96
2019-02-02	47145.79	41363.12	13.98
2019-02-03	11302.42	10102.41	11.88
2019-02-04	20981.08	19942.17	5.21
2019-02-05	8367.78	7067.78	18.39
2019-02-06	23686.76	21681.31	9.25
2019-02-07	6039.23	5239.11	15.27
2019-02-08	11186.27	10126.12	10.47
2019-02-09	21543.86	17421.12	23.67
2019-02-10	19659.91	15231.21	29.08
2019-02-11	23029.42	19579.42	17.62
2019-02-12	27847.98	21847.98	27.46
2019-02-13	45241.21	42131.21	7.38
2019-02-14	8472.69	6542.69	29.50
2019-02-15	42627.21	36227.21	17.67
2019-02-16	2514.03	1914.03	31.35
2019-02-17	19942.47	15942.47	25.09
2019-02-18	12496.78	9496.78	31.59
2019-02-19	24323.43	17323.43	40.41
2019-02-20	4279.68	3279.68	30.49
2019-02-21	15334.9	10314.9	48.67
2019-02-22	26819.8	23124.54	15.98
2019-02-23	19778.14	15778.06	25.35
2019-02-24	20702.01	17702.13	16.95
2019-02-25	12401.22	11401.65	8.77
2019-02-26	21081.33	19081.23	10.48
2019-02-27	13841.41	12341.86	12.15

2019-02-28	30674.99	21674.01	41.53
2018-03-01	12453.1	11453.17	8.73
2018-03-02	27616.04	20616.56	33.95
2018-03-03	11394.96	9394.12	21.30
2018-03-04	15620.74	12340.45	26.58
2018-03-05	14613.79	13113.19	11.44
2018-03-06	30017.93	27411.53	9.51
2018-03-07	23943.72	21253.29	12.66
2018-03-08	29258.55	26152.12	11.88
2018-03-09	22098.98	18991.22	16.36
2018-03-10	22945.1	19941.23	15.06
2018-03-11	45627.21	39613.09	15.18
2018-03-12	23688.28	21413.12	10.63
2018-03-13	25084.83	23218.13	8.04
2018-03-14	19116.11	15736.23	21.48
2018-03-15	16215.45	13901.32	16.65
2018-03-16	14803.21	13513.16	9.55
2018-03-17	15779.68	14149.78	11.52
2018-03-18	23468.14	21168.74	10.86
2018-03-19	48495.69	39131.11	23.93
2018-03-20	13576.89	11236.11	20.83
2018-03-21	32444.01	30251.06	7.25
2018-03-22	28078.21	24748.01	13.46
2018-03-23	53636.19	51071.16	5.02
2018-03-24	34444.08	32314.18	6.59
2018-03-25	25573.58	21573.2	18.54
2018-03-26	11960.36	9940.11	20.32
2018-03-27	15053.21	13512.12	11.41
2018-03-28	9521.81	8523.81	11.71
2018-03-29	34085.05	27231.77	25.17
2018-03-30	30773.49	28643.01	7.44
2018-03-31	17344.23	15189.54	14.19

Tabla 20 *Ficha de registros- rotación de mercancías (Post -Test)*

N ° DE FICHA DE REGISTRO:	4
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/04/2019-31/05/2019

Vari able	Indicador	Descripción	Técni ca	Med ida	Instrume nto	Formula
Logí stica	Porcentaje de rotación de mercancías	Movimiento de mercancías	Obser vació n	Por cient o	Ficha de observació n	$VP = \frac{VENTAS}{I.P.R \text{ OMEDIO}} * 100$

FECHA	VENTA	INVENTARIO PROMEDIO	INDICADOR
2019-04-01	13151.01	4611.00	2.85
2019-04-02	51021.01	7134.00	7.15
2019-04-03	13712.14	3450.00	3.97
2019-04-04	25131.81	5235.00	4.80
2019-04-05	10345.79	3523.00	2.94
2019-04-06	27431.12	4564.00	6.01
2019-04-07	6921.77	2431.00	2.85
2019-04-08	13315.65	4531.00	2.94
2019-04-09	24827.04	4270.00	5.81
2019-04-10	24231.12	3956.00	6.13
2019-04-11	27561.01	3534.00	7.80
2019-04-12	27847.98	5809.00	4.79
2019-04-13	13933.23	2650.00	5.26
2019-04-14	9841.23	2700.00	3.64
2019-04-15	45432.61	9923.00	4.58
2019-04-16	4241.45	2947.00	1.44
2019-04-17	21349.56	3987.00	5.35
2019-04-18	13926.78	4151.00	3.36
2019-04-19	29963.43	3850.00	7.78
2019-04-20	8292.68	3731.00	2.22
2019-04-21	19431.9	2679.00	7.25
2019-04-22	28919.8	3754.00	7.70
2019-04-23	19778.14	4216.00	4.69
2019-04-24	23328.78	2751.00	8.48
2019-04-25	14187.01	2970.00	4.78
2019-04-26	24251.11	3451.00	7.03
2019-04-27	16811.41	4150.00	4.05
2019-04-28	35141.09	4353.00	8.07
2019-04-29	12653.1	2310.00	5.48

2019-04-30	28616.04	2919.00	9.80
2019-05-01	14394.96	2950.00	4.88
2019-05-02	15920.74	3759.00	4.24
2019-05-03	15459.23	4332.00	3.57
2019-05-04	31917.93	4500.00	7.09
2019-05-05	23943.72	3701.00	6.47
2019-05-06	31258.55	3141.00	9.95
2019-05-07	22098.98	3198.00	6.91
2019-05-08	23846.05	3094.00	7.71
2019-05-09	48027.21	10602.00	4.53
2019-05-10	24365.23	2709.00	8.99
2019-05-11	25884.83	3160.00	8.19
2019-05-12	20451.34	3646.00	5.61
2019-05-13	16215.45	2752.00	5.89
2019-05-14	15543.21	3501.00	4.44
2019-05-15	15779.68	3605.00	4.38
2019-05-16	24323.14	2401.00	10.13
2019-05-17	48495.69	10455.00	4.64
2019-05-18	14351.23	4060.00	3.53
2019-05-19	33231.56	9682.00	3.43
2019-05-20	29921.98	4012.00	7.46

Tabla 21 *Ficha de registro- volumen de compras (Post Test)*

N ° DE FICHA DE REGISTRO:	5
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/04/2019-31/05/2019

Variable	Indicador	Descripción	Técnica	Medida	Instrumento	Formula
Logística	Porcentaje de volumen de compras	Volumen de compras	Observación	Porcentaje	Ficha de observación	$VP = \frac{(COMPRA/VENTAS) * 100}{1}$

FECHA	VENTA	COMPRAS	INDICADOR
2019-04-01	13151.01	3345.99	25.44
2019-04-02	51021.01	12345.45	24.20
2019-04-03	13712.14	4925.6	35.92
2019-04-04	25131.81	6521.45	25.95
2019-04-05	10345.79	3211.95	31.05
2019-04-06	27431.12	6787.69	24.74
2019-04-07	6921.77	1787.76	25.83
2019-04-08	13315.65	3323.23	24.96
2019-04-09	24827.04	6634.43	26.72
2019-04-10	24231.12	6214.98	25.65
2019-04-11	27561.01	6952.23	25.22
2019-04-12	27847.98	7445	26.73
2019-04-13	13933.23	3523.01	25.28
2019-04-14	9841.23	2345.56	23.83
2019-04-15	45432.61	12248.23	26.96
2019-04-16	4241.45	1315.51	31.02
2019-04-17	21349.56	5032.55	23.57
2019-04-18	13926.78	3456.2	24.82
2019-04-19	29963.43	6321.56	21.10
2019-04-20	8292.68	2823.23	34.04
2019-04-21	19431.9	5233.72	26.93
2019-04-22	28919.8	7704.95	26.64
2019-04-23	19778.14	4644.11	23.48
2019-04-24	23328.78	8954.43	38.38
2019-04-25	14187.01	3831.45	27.01
2019-04-26	24251.11	6521.33	26.89
2019-04-27	16811.41	4945.23	29.42

2019-04-28	35141.09	11238.75	31.98
2019-04-29	12653.1	4513.65	35.67
2019-04-30	28616.04	8323.23	29.09
2019-05-01	14394.96	2923.23	20.31
2019-05-02	15920.74	3923.45	24.64
2019-05-03	15459.23	3853.45	24.93
2019-05-04	31917.93	7134.48	22.35
2019-05-05	23943.72	5685.93	23.75
2019-05-06	31258.55	7967.64	25.49
2019-05-07	22098.98	5524.74	25.00
2019-05-08	23846.05	5936.28	24.89
2019-05-09	48027.21	12562.64	26.16
2019-05-10	24365.23	7522.07	30.87
2019-05-11	25884.83	7771.21	30.02
2019-05-12	20451.34	5579.03	27.28
2019-05-13	16215.45	4553.86	28.08
2019-05-14	15543.21	5700.8	36.68
2019-05-15	15779.68	3934.92	24.94
2019-05-16	24323.14	6467.04	26.59
2019-05-17	48495.69	11623.92	23.97
2019-05-18	14351.23	3994.22	27.83
2019-05-19	33231.56	12678.28	38.15
2019-05-20	29921.98	8119.55	27.14

Tabla 22 *Ficha de registro- Crecimiento de Ventas (Post Test)*

N ° DE FICHA DE REGISTRO:	6
Observadores:	Cáceres Guerra Walther y Coronel Yoel
Institución donde se investiga:	Villa chicken S.A.C
Indicador observado:	Av. Francisco Javier Mariátegui 912, Jesús María
Periodo de observación:	01/04/2019-31/05/2019

Var iabl e	Indicador	Descripci ón	Técni ca	Me dida	Instrume nto	Formula
Log ístic a	Porcentaje de crecimiento de ventas	Crecimien to de ventas	Obser vació n	Por cien to	Ficha de observaci ón	$VP = \frac{(VENTAS\ 2019 - VENTAS2018)}{VENTAS\ S2018} * 100$

FECHA	VENTAS2019	Ventas2018	INDICADOR
2019-04-01	13151.01	11121.01	18.25
2019-04-02	51021.01	41465.56	23.04
2019-04-03	13712.14	10952.41	25.20
2019-04-04	25131.81	20942.17	20.01
2019-04-05	10345.79	8067.23	28.24
2019-04-06	27431.12	21681.31	26.52
2019-04-07	6921.77	5423.67	27.62
2019-04-08	13315.65	11178.54	19.12
2019-04-09	24827.04	19786.89	25.47
2019-04-10	24231.12	16934.03	43.09
2019-04-11	27561.01	21452.42	28.48
2019-04-12	27847.98	21847.98	27.46
2019-04-13	13933.23	11231.98	24.05
2019-04-14	9841.23	7023.24	40.12
2019-04-15	45432.61	36227.21	25.41
2019-04-16	4241.45	3076.03	37.89
2019-04-17	21349.56	15942.47	33.92
2019-04-18	13926.78	9496.78	46.65
2019-04-19	29963.43	19331.56	55.00
2019-04-20	8292.68	5738.34	44.51
2019-04-21	19431.9	12456.45	56.00
2019-04-22	28919.8	23124.54	25.06
2019-04-23	19778.14	15778.06	25.35
2019-04-24	23328.78	17802.23	31.04
2019-04-25	14187.01	11634.45	21.94
2019-04-26	24251.11	19234.34	26.08
2019-04-27	16811.41	13341.23	26.01

2019-04-28	35141.09	23674.01	48.44
2019-04-29	12653.1	11453.17	10.48
2019-04-30	28616.04	20616.56	38.80
2019-05-01	14394.96	11094.12	29.75
2019-05-02	15920.74	12340.45	29.01
2019-05-03	15459.23	13113.19	17.89
2019-05-04	31917.93	27411.53	16.44
2019-05-05	23943.72	20453.3	17.07
2019-05-06	31258.55	26152.12	19.53
2019-05-07	22098.98	18991.22	16.36
2019-05-08	23846.05	19941.23	19.58
2019-05-09	48027.21	39613.09	21.24
2019-05-10	24365.23	21413.12	13.79
2019-05-11	25884.83	23218.13	11.49
2019-05-12	20451.34	15736.23	29.96
2019-05-13	16215.45	13401.32	21.00
2019-05-14	15543.21	13513.16	15.02
2019-05-15	15779.68	14149.78	11.52
2019-05-16	24323.14	21168.74	14.90
2019-05-17	48495.69	35631.11	36.10
2019-05-18	14351.23	11236.11	27.72
2019-05-19	33231.56	30251.06	9.85
2019-05-20	29921.98	25212.56	18.68

Anexo 3: Desarrollo

1. Comprensión del Negocio

1.1 Determinar los objetivos de Negocio

El objetivo de la minería de datos que se va aplicar en este proyecto de investigación son las predicciones más precisas posibles a partir de los datos que se dispone la Base de datos de Villa Chicken. El objetivo es mantener el stock controlado para evitar unos escases de abastecimiento para reducir las ventas perdidas. Ya que al no entregar el pedido solicitado al cliente, se convierte en un ingreso que deja de percibir la organización.

Objetivos de negocio

- Predecir la Cantidad de productos que se venderá una determina fecha Festiva (**San Valentín, Día de la Mujer, Día del Trabajador, Día de la Madre, Día de la papa, y Día del pollo**) del año 2019.
- Hacer Predicciones sobre cantidad que se venderá en un determinado mes de mayo y junio con respecto al producto POLLO S/M GRANEL B4 3, del año 2019.

1.2 Evaluación de la situación

Se cuenta con un servidor Proliant DL580 G5, con un sistema operativo Windows server 2012 R2, con un gestor de base de datos Microsoft Sql Server 2012 Enterprise que contiene información del proceso logístico desde el año 2017, se puede afirmar que se dispone de una cantidad de datos para solucionar la problemática.

Para implementar el sistema de información, se propone los siguientes requerimientos de Software.

- Servidor web Apache (Xampp v3.3.3)
- Editor de texto Sublime Text v3
- Extensiones dll para conectar PHP 7.2 con Sql Server 2014
- Librería PHP Machine Learning

Evaluación inicial de Herramientas y técnicas

La herramienta que se va utilizar para llevar el proyecto de minería de datos es la librería PHP-ML, esta herramienta se acopla a la metodología que estamos empleando, La ventaja de esta herramienta es que no necesitamos importar un archivo CSV para realizar el proceso de minería de datos. Solo necesitamos incluir los datos en el script del PHP para el procesamiento.

En cuanto las técnicas que se va implementar para la extracción de conocimiento, **PHP-ML** proporciona las siguientes técnicas:

Predicción:

- Clasificación
- Regresión

La técnica de predicción que se utilizara en este proyecto de investigación es Regresión Lineal, Este modelo consiste en utilizar el método de mínimos cuadrados para aproximar una solución

2. Comprensión de los Datos

2.1 Recolectar los Datos Iniciales

Los Datos utilizados comprenden de todos los movimientos de rotación de inventarios histórico (POLLO S/M GRANEL B4, Papa Pre frita AVIKO Caja x 10 Kg y Churrasco (200gr)) de los años 2016 al 2018.

	IDHISTORICO	IDTIPO	IDINVENTARIO	FESTIVIDAD	Codigo	INVENTARIO	Año	CANTIDAD_VENDIDA
1	1	1	1	SAN VALENTIN	2203017	POLLO S/M GRANEL B4	2016	450.00
2	2	1	1	SAN VALENTIN	2203017	POLLO S/M GRANEL B4	2017	478.00
3	3	1	1	SAN VALENTIN	2203017	POLLO S/M GRANEL B4	2018	498.00
4	4	1	2	SAN VALENTIN	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2016	423.00
5	5	1	2	SAN VALENTIN	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2017	471.00
6	6	1	2	SAN VALENTIN	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2018	480.00
7	7	1	3	SAN VALENTIN	2204008	Churrasco (200gr)	2016	485.00
8	8	1	3	SAN VALENTIN	2204008	Churrasco (200gr)	2017	498.00
9	9	1	3	SAN VALENTIN	2204008	Churrasco (200gr)	2018	512.00
10	10	2	1	DÍA DE LA MUJER	2203017	POLLO S/M GRANEL B4	2016	390.00
11	11	2	1	DÍA DE LA MUJER	2203017	POLLO S/M GRANEL B4	2017	398.00
12	12	2	1	DÍA DE LA MUJER	2203017	POLLO S/M GRANEL B4	2018	411.00
13	13	2	2	DÍA DE LA MUJER	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2016	205.00
14	14	2	2	DÍA DE LA MUJER	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2017	236.00
15	15	2	2	DÍA DE LA MUJER	1203014	Papa Pre frita AVIKO Caja x 10 Kg	2018	240.00
16	16	2	3	DÍA DE LA MUJER	2204008	Churrasco (200gr)	2016	467.00
17	17	2	3	DÍA DE LA MUJER	2204008	Churrasco (200gr)	2017	465.00

Figura 18: Recolección de Datos (Movimiento de inventario por Fecha Festiva periodo 2016 al 2018)- Fase Inicial

La función nos devolvió un total de 54 Registros necesarios para entrenar el clasificador.

A continuación, listamos los datos Adquiridos:

- **IdHistorico**

Cada Transacción está asociado a un solo código único.

- **IdTipo**

Identificador que está asociado al tipo de Fecha Festiva.

- **IdInventario**

Identificador único que está asociado a cada inventario.

- **Festividad**

Se menciona que días caen una fecha festiva (**San Valentín, Día de la Mujer, Día del Trabajador, Día de la Madre, Día de la papa, y Día del pollo**)

- **Código**

Identificador único asociado a cada inventario.

- **Inventario**

Nombre del inventario (POLLO S/M GRANEL B4, Papa Pre frita AVIKO Caja x 10 Kg y Churrasco (200gr)).

- **Año**

Cada transacción está asociado a un periodo de 2016 al 2018.

- **Cantidad Vendida**

Cada transacción está asociada a cantidad vendida.

	IdHistorico	IdInventario	FECHA	Cantidad	IdInventario	Codigo	Descripcion	Estado
1	122	1	2016-05-01	355.00	1	2203017	POLLO S/M GRANEL B4	1
2	123	1	2016-05-02	323.00	1	2203017	POLLO S/M GRANEL B4	1
3	124	1	2016-05-03	341.00	1	2203017	POLLO S/M GRANEL B4	1
4	125	1	2016-05-04	324.00	1	2203017	POLLO S/M GRANEL B4	1
5	126	1	2016-05-05	289.00	1	2203017	POLLO S/M GRANEL B4	1
6	127	1	2016-05-06	286.00	1	2203017	POLLO S/M GRANEL B4	1
7	128	1	2016-05-07	324.00	1	2203017	POLLO S/M GRANEL B4	1
8	129	1	2016-05-08	383.00	1	2203017	POLLO S/M GRANEL B4	1
9	130	1	2016-05-09	323.00	1	2203017	POLLO S/M GRANEL B4	1
10	131	1	2016-05-10	291.00	1	2203017	POLLO S/M GRANEL B4	1
11	132	1	2016-05-11	333.00	1	2203017	POLLO S/M GRANEL B4	1
12	133	1	2016-05-12	321.00	1	2203017	POLLO S/M GRANEL B4	1
13	134	1	2016-05-13	312.00	1	2203017	POLLO S/M GRANEL B4	1
14	135	1	2016-05-14	296.00	1	2203017	POLLO S/M GRANEL B4	1
15	136	1	2016-05-15	315.00	1	2203017	POLLO S/M GRANEL B4	1
16	137	1	2016-05-16	299.00	1	2203017	POLLO S/M GRANEL B4	1
17	138	1	2016-05-17	325.00	1	2203017	POLLO S/M GRANEL B4	1

Figura 19 Recolección de Datos (Movimiento de inventario Mensual(Mayo-Junio) Festiva periodo 2016 al 2018)- Fase Inicial

La función nos devolvió un total de 183 Registros necesarios para entrenar el clasificador.

A continuación, listamos los datos Adquiridos:

- **IdHistorico**

Cada Transacción está asociado a un solo código único.

- **IdTipo**

Identificador que está asociado a cada transacción diaria.

- **IdInventario**

Identificador único que está asociado a cada inventario.

- **Fecha**

Transacciones diarias que corresponde a los meses de mayo-junio del año 2016-2018

- **Cantidad**

Cada transacción está asociada a cantidad vendida.

- **Código**

Identificador único asociado a cada inventario.

- **Inventario**

Nombre del inventario (POLLO S/M GRANEL B4, Papa Pre frita AVIKO Caja x 10 Kg y Churrasco (200gr)).

3. Preparación de los Datos

3.1 Datos Seleccionados

En esta fase consiste en preparar los datos obtenidos para adecuarlos a las técnicas de minería de datos.

Con la extracción de datos obtenidas desde la Base de datos de Villa Chicken, se obtuvieron diferentes tipos de datos, para ello se seleccionó los atributos necesarios para el correcto funcionamiento del modelo.

```

/* SCRIPT PARA ENTRENAR LAS ROTACIONES DE INVENTARIOS POR FECHA FESTIVAS */
select I.Codigo,T.IdTipo,H.CANTIDAD from HISTORICO_MERCANCIA_FESTIVIDAD H
INNER JOIN Inventario I ON I.IdInventario=H.idInventario
inner join Tipo_Festividad t on T.IdTipo=h.IDTIPO

/* SCRIPT PARA ENTRENAR LAS ROTACION DE MERCANCIAS MENSUALES (Mayo-Junio) */
select H.FECHA,I.CODIGO,H.CANTIDAD from HISTORICO_MERCANCIA_FECHA H
INNER JOIN Inventario I ON I.IdInventario=H.idInventario
where MONTH(H.FECHA) in (5,6)

```

Figura 20 Script Sql Rotación de mercancías - Preparación de los Datos

3.2 Estructuración de los datos

Para crear los modelos con respecto a series del tiempo, los atributos utilizados son identificados de la siguiente manera:

- **Código**
Identificador único asociado a cada inventario.
- **IdTipo**
Identificador asociado al nombre de Fecha Festiva
- **Año**
Representada por cada Año
- **Cantidad**
El atributo Cantidad, representa el objetivo a predecir.

En esta fase preparamos los datos de la siguiente manera.

Tabla 23 Serie de Tiempo Rotación de Mercancías(POLLO S/M GRANEL B4) Anuales por Fecha Festiva - Preparación de los Datos

POLLO S/M GRANEL B4	2016	2017	2018
SAN VALENTIN	450	478	498
DÍA DE LA MUJER	390	398	411
DÍA DEL TRABAJADOR	355	376	391
DÍA DE LA MADRE	383	401	434
DÍA DE LA PAPA	489	501	538
DÍA DEL POLLO A LA BRASA	521	550	574

Tabla 24 *Serie de Tiempo Rotación de Mercancías(Papa Pre frita AVIKO Caja x 10 Kg) Anuales por Fecha Festiva - Preparación de los Datos*

Papa Pre frita AVIKO Caja x 10 Kg	2016	2017	2018
SAN VALENTIN	423	471	480
DÍA DE LA MUJER	205	236	240
DÍA DEL TRABAJADOR	321	343	398
DÍA DE LA MADRE	346	378	399
DÍA DE LA PAPA	564	576	592
DÍA DEL POLLO A LA BRASA	592	616	642

Tabla 25 *Serie de Tiempo Rotación de Mercancías(Churrasco (200gr)) Anuales por Fecha Festiva - Preparación de los Datos*

Churrasco (200gr)	2016	2017	2018
SAN VALENTIN	485	498	512
DÍA DE LA MUJER	467	465	472
DÍA DEL TRABAJADOR	405	425	453
DÍA DE LA MADRE	397	422	439
DÍA DE LA PAPA	456	489	494
DÍA DEL POLLO A LA BRASA	472	483	492

Tabla 26 *Serie de Tiempo Rotación de Mercancías POLLO S/M GRANEL B4)- Mensual (Mayo) - Preparación de los Datos*

POLLO S/M GRANEL B4	MAYO		
Día	2016	2017	2018
1	355	376	391
2	323	342	376
3	341	357	353
4	324	343	372
5	289	345	381
6	286	302	325
7	324	331	364
8	383	402	467
9	323	332	361
10	291	317	342
11	333	354	370
12	321	344	382
13	312	317	434
14	296	401	442
15	315	351	392
16	299	332	363
17	325	342	376

18	315	345	361
19	291	316	337
20	356	347	363
21	313	349	362
22	246	298	325
23	341	357	374
24	258	298	324
25	351	362	362
26	324	357	378
27	315	364	342
28	249	297	304
29	309	341	371
30	489	501	538
31	352	343	352

Tabla 27 *Serie de Tiempo Rotación de Mercancías POLLO S/M GRANEL B4)- Mensual (Junio) - Preparación de los Datos*

POLLO S/M GRANEL B4	JUNIO		
Día	2016	2017	2018
1	331	343	361
2	234	252	286
3	268	296	325
4	307	323	356
5	265	298	321
6	307	323	339
7	351	334	348
8	297	347	361
9	301	365	378
10	235	293	302
11	309	323	364
12	234	253	293
13	346	375	391
14	345	361	382
15	237	277	317
16	356	374	392
17	239	283	302
18	321	354	375
19	235	264	276
20	322	335	342
21	314	351	378
22	297	354	373
23	305	341	364

24	307	356	368
25	291	324	343
26	301	333	351
27	321	346	362
28	312	332	371
29	322	343	346
30	298	354	367

4. Modelamiento

En esta fase de la metodología se escogerá la técnica de data mining para poder predecir. Ya lo mencionado en la fase 1, la técnica que se utilizará en este proyecto de investigación es Regresión Lineal simple, Este modelo consiste en utilizar el método de mínimos cuadrados para aproximar una solución

En la Figura N° 23, se aplica el entrenamiento para obtener el pronóstico de cantidad de rotación de mercancías por fecha festiva.

```

1  <?php
2  // API MACHINE LEARNING REGRESSION LINEAL
3  require_once __DIR__ . '/../vendor/autoload.php';
4  use Phpml\regression\LeastSquares;
5  use Phpml\Metric\ClassificationReport;
6  use Phpml\FeatureExtraction\TokenCountVectorizer;
7  use Phpml\Tokenization\WordTokenizer;
8  $samples = [[2016],[2017],[2018]];
9  $targets = [450.00,478.00,498.00,];
10 $regression = new LeastSquares();
11 $regression->train($samples, $targets);
12 echo "2203017"."-1-".$regression->predict([2019])."</br>";
13
14 $samples = [[2016],[2017],[2018]];
15 $targets = [390.00,398.00,411.00,];
16 $regression = new LeastSquares();
17 $regression->train($samples, $targets);
18 echo "2203017"."-2-".$regression->predict([2019])."</br>";
19
20 $samples = [[2016],[2017],[2018]];
21 $targets = [355.00,376.00,391.00,];
22 $regression = new LeastSquares();
23 $regression->train($samples, $targets);
24 echo "2203017"."-3-".$regression->predict([2019])."</br>";
25
26 $samples = [[2016],[2017],[2018]];
27 $targets = [383.00,401.00,434.00,];
28 $regression = new LeastSquares();
29 $regression->train($samples, $targets);
30 echo "2203017"."-4-".$regression->predict([2019])."</br>";
31
32

```

Figura 21: Modelamiento del clasificador desde el Sistema de Predicción Logístico- Valores de entrenamiento

PHP-ML Realiza el proceso de entrenamiento con esta función para predecir resultados.

5. Evaluación

En esta fase se evalúa los modelos generados desde el punto de vista de los objetivos de negocio.

Evaluación de Resultados

- Modelo del objetivo 1

Tabla 28 *Evaluación de Resultados Objetivo 1 desde el Sistema de Predicción Logístico*

2019	POLLO S/M GRANEL B4	Papa Pre frita AVIKO Caja x 10 Kg	Churrasco (200gr)
SAN VALENTIN	547	543	538
DÍA DE LA MUJER	431	279	475
DÍA DEL TRABAJADOR	428	469	499
DÍA DE LA MADRE	482	453	482
DÍA DE LA PAPA	558	605	517
DÍA DEL POLLO A LA BRASA	601	666	502

En este modelo es aceptable ya que podemos hacer precisiones de la cantidad que se venderá en una determina fecha festiva. (**San Valentín, Día de la Mujer, Día del Trabajador, Día de la Madre, Día de la papa, y Día del pollo**)

- Modelo del objetivo 2

Tabla 29 *Evaluación de Resultados Objetivo 2 desde el Sistema de Predicción Logístico*

POLLO S/M GRANEL B4	MAYO	JUNIO
Día	2019	2019
1	409	374
2	400	309
3	362	353
4	394	377
5	430	350
6	343	354
7	379	341
8	501	398
9	376	425
10	367	343
11	389	387
12	409	319
13	476	415

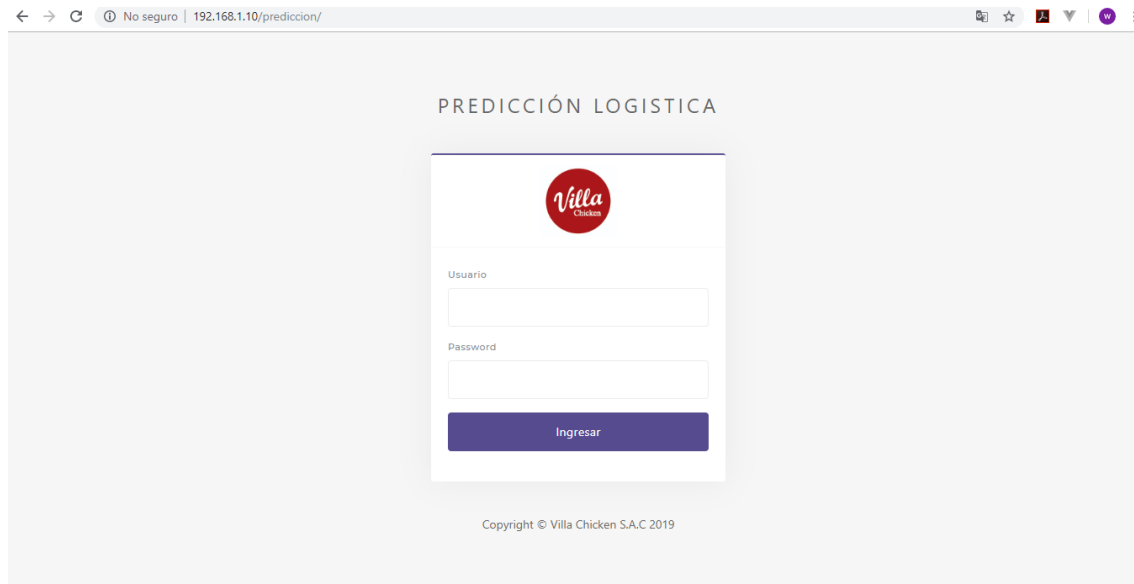
14	525	399
15	429	356
16	395	417
17	398	337
18	386	403
19	360	299
20	362	353
21	390	411
22	368	417
23	390	395
24	359	404
25	369	371
26	406	378
27	367	383
28	338	397
29	402	361
30	558	408
31	349	0

En este segundo modelo es factible yaqué se pueden hacer predicciones sobre la cantidad que se va vender en un determinado mes con respecto a un producto seleccionado (POLLO S/M GRANEL B4).

1. Implementación

Diseño:

Interfaz de usuario, Sistema de Predicción Logística



PREDICCIÓN LOGÍSTICA

Villa Chicken

Usuario

Password

Ingresar

Copyright © Villa Chicken S.A.C 2019

Figura 22 : Sistema de predicción Logístico- Inicio Sesión

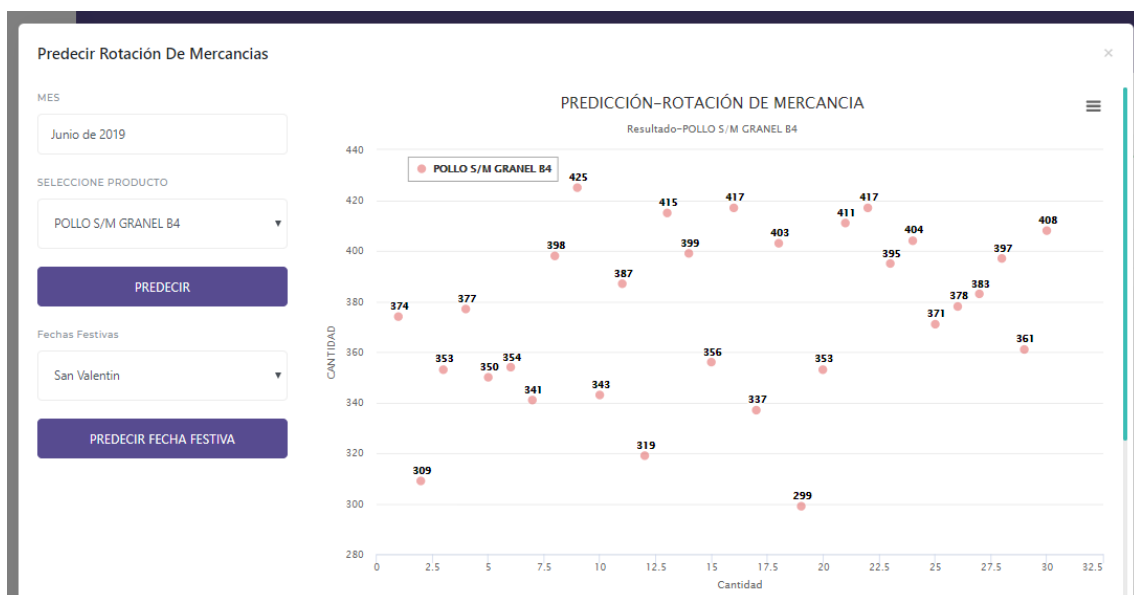


Figura 23: Sistema de predicción Logístico – Predicción cantidad de rotación POLLO S/M GGRANEL B4- Mensual

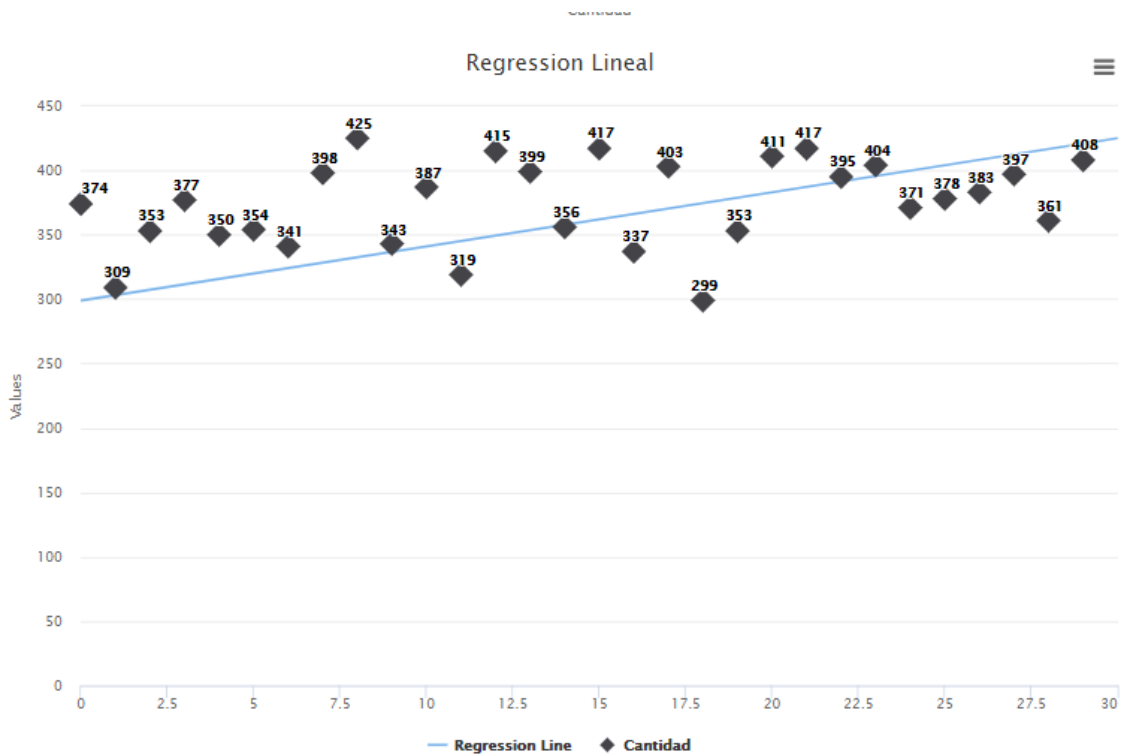


Figura 24: Sistema de predicción Logístico – Predicción cantidad rotación POLLO S/M GGRANEL B4- MENSUAL – Regresión Lineal

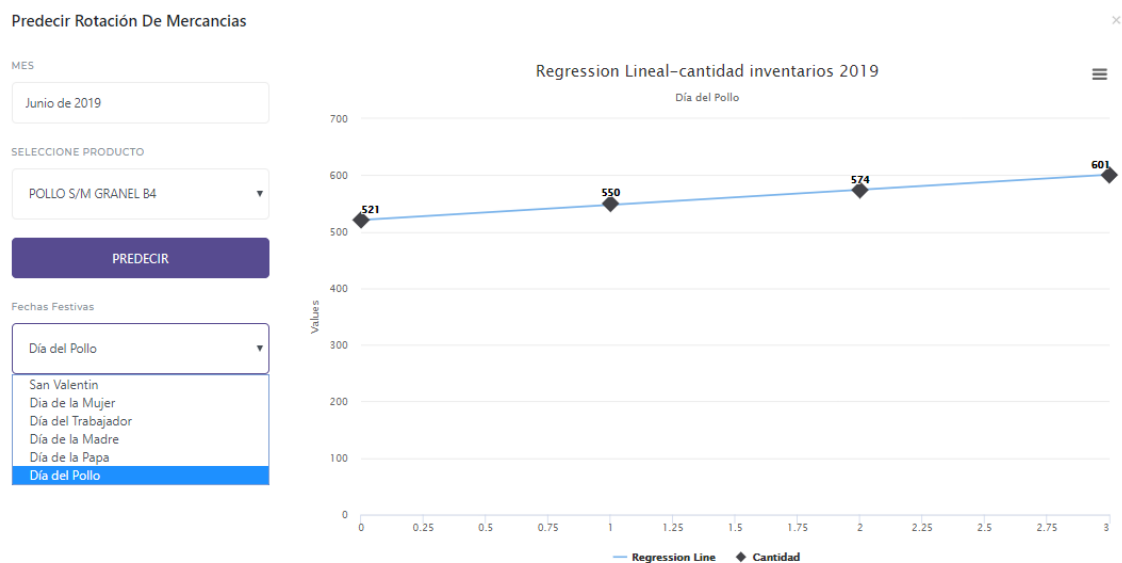


Figura 25: Sistema de predicción Logístico - Predicción cantidad rotación POLLO S/M GGRANEL B4 2019 – Día del Pollo-Regresión Lineal

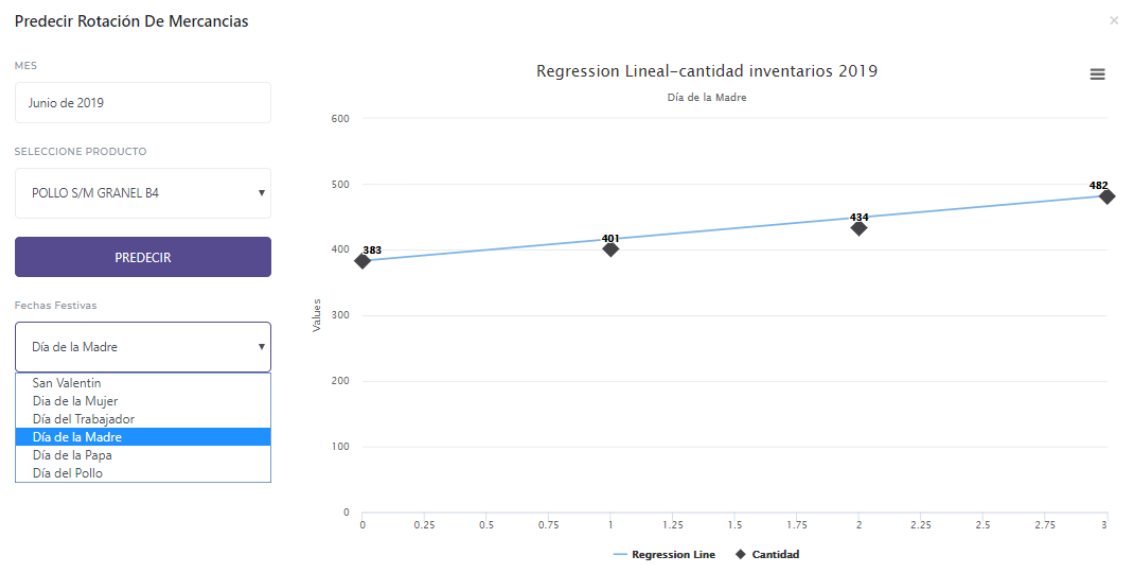


Figura 26: Sistema de predicción Logístico – Predicción de cantidad de rotación POLLO S/M GGRANEL B4 2019 FECHA FESTIVO -DÍA DE LA MADRE- Regresión Lineal

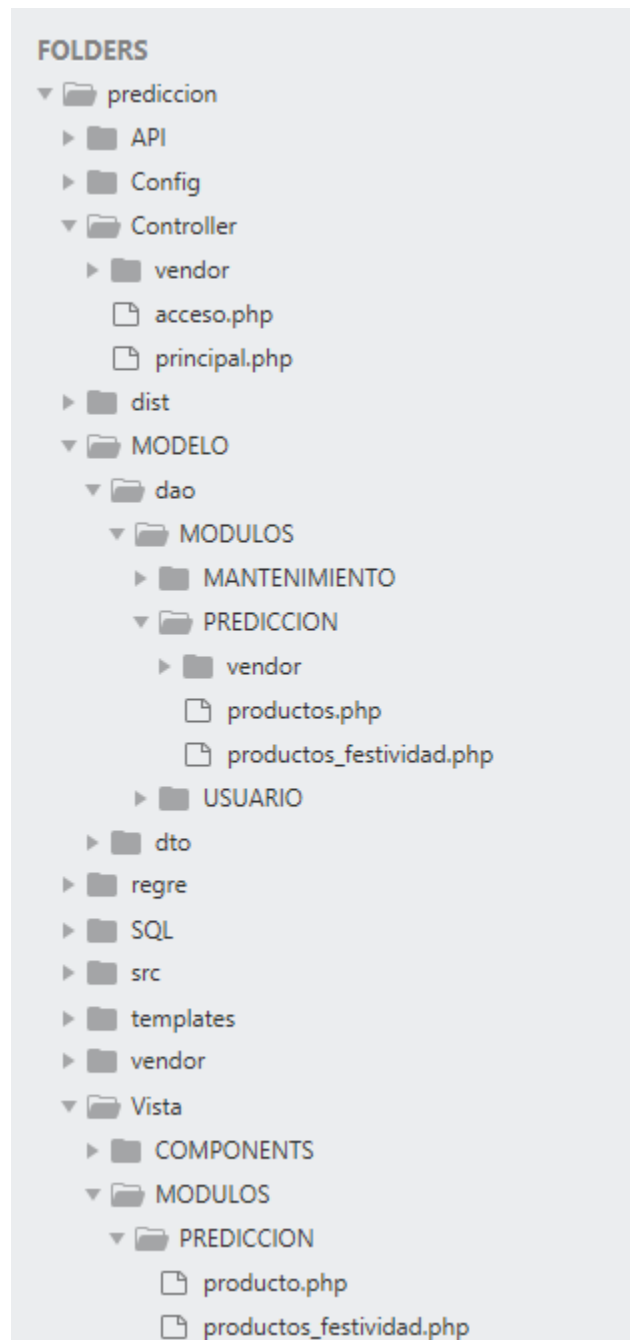


Figura 27: Estructura del aplicativo web- Sistema predicción logística.

```

<?php
// API MACHINE LEARNING
require_once __DIR__ . '/vendor/autoload.php';
session_name('PREDICCION_LOGISTICA');
session_start();
date_default_timezone_set('America/Lima');
include '../sql/conexion_api.php';
$instruccion=isset($_GET['p'])?$_GET['p']:'';
//LOGICA
include '../MODELO/dao/MODULOS/PREDICCION/productos.php';
include '../MODELO/dao/MODULOS/PREDICCION/productos_festividad.php';
//VISTAS
include '../VISTA/MODULOS/PREDICCION/producto.php';
include '../VISTA/MODULOS/PREDICCION/productos_festividad.php';
$mod_prediccion_producto = new MODELO_PREDICCION_PRODUCTO();
$mod_prediccion_producto_festividad = new MODELO_PREDICCION_PRODUCTO_FESTIVIDAD();
$instruccion=isset($_GET['p'])?$_GET['p']:'';
/*PREDICCION CANTIDAD PRODUCTO MENSUAL */
if($instruccion=='PREDECIR_PRODUCTO'){
    if(isset($_POST['MES']) && isset($_POST['PRODUCTO']))
    {
        $porciones = explode("-",$_POST['MES']);
        $mod_prediccion_producto->GET_RESULTADO_PREDICCION_MENSUAL($_POST['PRODUCTO'],$porciones[1],$_POST['NOMBRE']);
    }else{
    }
}
}
/*PREDICCION PRODUCTO POR FESTIVIDAD */
else if($instruccion=='PREDECIR_FECHA_FESTIVA_PRODUCTO'){
    if(isset($_POST['FECHA']) && isset($_POST['FECHA_NOMBRE']) && isset($_POST['PRODUCTO']) )
    {
        $mod_prediccion_producto_festividad->GET_RESULTADO_PREDICCION_FESTIVIDAD($_POST['PRODUCTO'],$_POST['FECHA'],$_POST['FECHA_NOMBRE']);
    }else{
    }
}
}

```

Figura 28: Sistema de predicción Logística – Capa Contoller

```

<?php
use Phpml\regression\LeastSquares;
use Phpml\Metric\ClassificationReport;
use Phpml\FeatureExtraction\TokenCountVectorizer;
use Phpml\Tokenization\WordTokenizer;
class MODELO_PREDICCION_PRODUCTO_FESTIVIDAD extends conexion_api{
    public $conexion;
    function __construct()
    {
        parent::__construct();
        $cn= new conexion_api();
        $this->conexion=$cn->getConnection();
    }

    function GET_RESULTADO_PREDICCION_FESTIVIDAD($producto,$fecha,$nombre){
        $salida="";
        $query="
declare @codigo varchar(50)
declare @tipo_festividad int

set @codigo=?
set @tipo_festividad=?

select T.Descripcion AS FESTIVIDAD,Codigo,I.DESCRIPCION,Año,cantidad from HISTORICO_MERCANCIA_FESTIVIDAD PF
INNER JOIN Inventario I ON I.IdInventario=PF.IdInventario
inner join Tipo_Festividad t on T.IdTipo=PF.IDTIPO
where Codigo=@codigo and T.IdTipo=@tipo_festividad
union all
SELECT T.Descripcion AS FESTIVIDAD,Codigo,I.DESCRIPCION,'2019' as Año ,cantidad FROM Prediccion_Festividad PF
INNER JOIN Inventario I ON I.IdInventario=PF.IdInventario
inner join Tipo_Festividad t on T.IdTipo=PF.IDTIPO
where Codigo=@codigo and t.IdTipo=@tipo_festividad
";

        $array=array();
        $ejecutar=$this->conexion->prepare($query,array(PDO::ATTR_CURSOR => PDO::CURSOR_SCROLL));
        $ejecutar->bindParam(1,$producto, PDO::PARAM_STR);
    }
}

```

Figura 29: Sistema de Predicción Logística – Capa Model

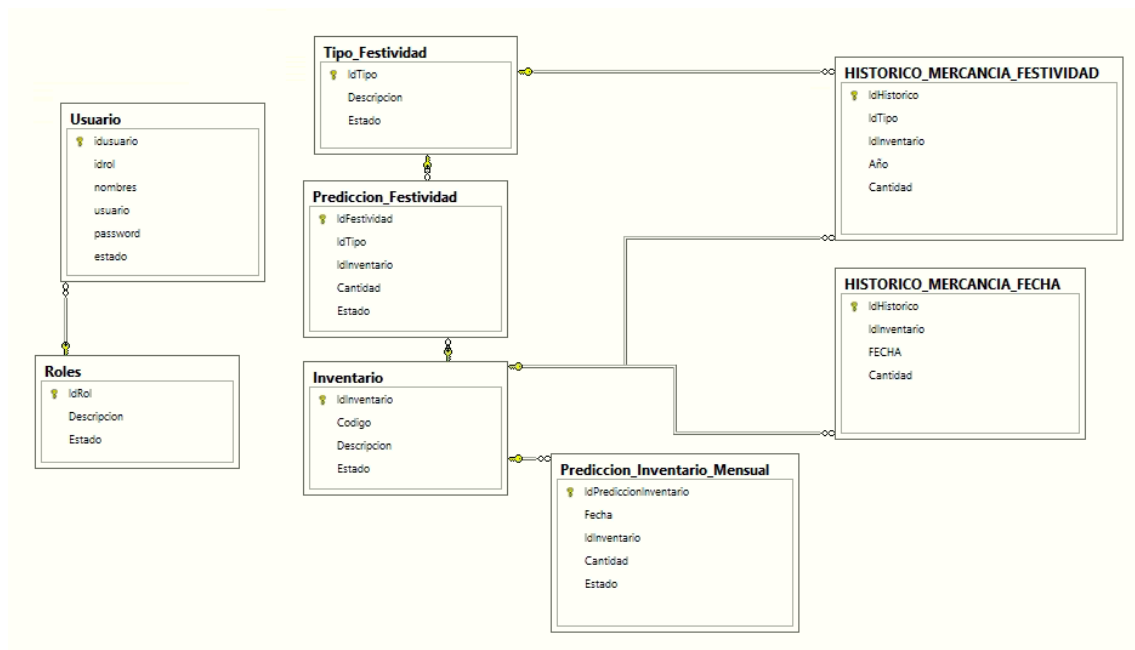


Figura 30 Modelo Físico de la Base De datos

Anexo 4: Permiso de uso de datos de institución

CONSTANCIA DE USO DE DATOS

Jefe de Tecnología de Información Villa Chicken S.A.C.

Ing. Alejandro J. Wu Wong

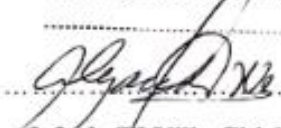
Deja constancia:

Que el señor Yoel Kereli Coronel Araujo identificados con DNI N° 72024497, viene realizando un proyecto de investigación en el área de logística de Villa Chicken S.A.C., para lo cual se le brindo dos muestras de datos para tal investigación. Así mismo aseguramos que el señor Yoel Kereli Coronel Araujo estuvo implementado un aplicativo en nuestra institución.

Se expide tal documento para los fines que el interesado estime convenientemente.

Lima, 13 de octubre 2020

Villa Chicken S.A.C.


.....
Jefe de TI Villa Chicken